AD734155

Instructional Strategies:
Multivariable Studies of Psychological Processes Related to Instruction

Francis J. Di Vesta   Nicholas M. Sanders
Charles B. Schultz   Paul D. Weener

Annual Report

July, 1971

Part II

Advanced Research Projects Agency
ARPA Order No. 1269
Monitored by
Office of Naval Research
Contract Number N00014-67-A-0385-0006

Department of Educational Psychology
College of Education
The Pennsylvania State University

DDC
RECEIVED
DEC 23 1971
B

PART #1 -AD 730032

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Reproduced by
NATIONAL TECHNICAL INFORMATION SERVICE
Springfield, Va. 22151

DISCLAIMER NOTICE

THIS DOCUMENT IS THE BEST QUALITY AVAILABLE.

COPY FURNISHED CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

ANNUAL REPORT

Part II

January 1, 1970 - June 30, 1971

ARPA Order Number:	1269
Program Code Number:	60949
Name of Contractor:	Pennsylvania State University
	University Park, Pennsylvania 16802
Effective Date of Contract:	1968 September 1
Contract Expiration Date:	1971 June 30
Contract Number:	N00014-67-A-0385-0006
Principal Investigator	Francis J. Di Vesta
and Phone Number:	865-8303
Title of Work:	Instructional Strategies

Sponsored by
Advanced Research Projects Agency

ARPA Order No. 1269

TABLE OF CONTENTS

Section I

Project Ikon:

Studies of Imagery

Summary

Project Ikon: Studies of Imagery

Francis J. Di Vesta

In collaboration with

G. Susan Gray, Gary Ingersoll, Steven Ross, and Phyllis Sunshine

Technical Problem

This was a program of research consisting of seven studies in which were investigated the interactions between imagery-ability and experimental treatments that parallel instructional procedures. The general orientation was to extend the studies of imaginal and verbal coding systems of learning, operationally defined in terms of stimulus attributes (i.e., concreteness and abstractness) or instructional sets (i.e., imaginal vs. verbal processing) to include individual differences in symbolic habits. Thus, all of the present studies took into account the learner's ability to use the imaginal strategy, which was called imagery-ability, as the primary individual difference variable. Underlying this series of systematic studies was the reasoning that imagery-ability (i.e., high- and low-imagery) could be related to the attributes of stimuli (i.e., rated-imagery), or to the processes employed by the learner in acquiring, storing, and retrieving information (i.e., imaginal vs. verbal processing).

General Methodology

Initially, a series of three experiments were conducted in an attempt to replicate those conducted by Stewart (1965). In two of the studies the task materials were presented pictorially and verbally in contrastive treatments. In one of these two studies we examined the ability of high- and low-imagers to transfer material, learned via pictures and verbalizations, to new situations while in the other we examined the learner's ability to recall material presented via the two methods. In the third replication study, the influence of concrete and abstract verbal stimuli (rather than pictures and words) as variables affecting rate of acquisition and recall was investigated. A factor analytic study attempted to clarify the relationship between subjective reports and objective tests of imagery ability. In a fifth study we investigated, experimentally, the interaction between pictorial and verbal contexts of material to be learned on subsequent ability to transfer to new situations by high- and low-imagers. A sixth study was designed to examine the effect of imagery-ability in acquiring information from tasks in which the noun-adjective relationship in paired-associates was varied according to relationship between stimuli and responses (i.e., noun-adjective vs. adjective-noun sequence) and concreteness of the stimulus or response. In the seventh study the effectiveness of verbal and imaginal processing by high- and low-imagers was investigated.

Technical Results

In general, the results provided support for Paivio's (1970) two-stage model of associative learning as follows: The meanings of concrete words are learned via direct experience and intraverbal experience. Accordingly, they evoke both images and verbal responses. The meanings of abstract words, on the other hand, are learned primarily by association with other words, and, hence, elicit primarily verbal responses. Thus, imaginal processing is more effective for the processing of concrete words and verbal processing is more effective for the processing of abstract words.

Another conclusion, from these investigations, was that imagery-ability, as measured in these studies, reflects the ability of the learner to process the information by verbal or by imaginal strategies. There was little or no evidence that imagery ability reflected a sensitivity on the part of the learner to profit from pictorial or verbal stimuli, _per se_, as suggested by Stewart's (1965) studies. Nevertheless, the acquisition of all learners was facilitated more by pictorial than by verbal presentations. An interesting adjunct to this conclusion was that the picture-word order was always more favorable for learning, transfer, and recall than was the word-picture order. Furthermore, over several trials, varying the modality (e.g., presentation of a picture, then its verbal label, and then the picture again) facilitated recall to a greater extent than retaining the same modality (e.g., presentation of a picture) on all trials. Finally, the factor analytic study indicated that the verbal-ability and imagery-ability factors were orthogonal rather than bi-polar. Imagery tests based on subjective reports of ability to use imaginal processing was related primarily to social desirability rather than to objective tests of imagery ability. This latter finding undoubtedly accounts for some of the failures to identify relationships between individual differences in imagery-ability and ability of the person to profit from pictorial versus verbal materials either in terms of rate of acquisition or in terms of the facilitation of memory, but is not an exclusive reason.

Educational Implications

The implications of the results of these studies are, in all cases, more or less self-evident, if we are permitted to extrapolate from experimental to classroom settings. Thus, to mention a few implications: Imagery-ability can be measured from tests involving manipulation of objects in space. The acquisition of new concepts can be facilitated by providing direct experience prior to the provision of a label. Variations in the modality of stimuli tends to facilitate the recall of the material learned. Finally, imagery-ability means just that; high-imagers do use imaginal processing more effectively than do low-imagers. Conversely, low-imagers are severely handicapped when they are forced to employ imaginal strategies rather than verbal strategies. This finding implies that in adapting to this individual difference, high-imagers can be taught by methods which capitalize on imaginal processing (e.g., instruction which involves graphic displays) while low-imagers might be taught by methods which employ verbal processing (e.g., instruction via lecture methods).

Implications for Further Research

These studies indicate that the study of aptitude by treatment interactions with imagery would be most fruitful if the research strategy were to emphasize the nature of imaginal processing. Certainly, more needs to be known about the effects of the pictorial-labeling sequence in instructional strategy. However, as important is an understanding of the differences between verbal and imaginal processing. It would appear that the techniques employed in the present studies, with but some slight modification such as time-sampling could provide further insights into the way materials are transformed by the subject.

References

Paivio, A. On the functional significance of imagery. Psychological Bulletin, 1970, 6, 385-392.

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, University of Toronto, 1965.

The Recognition and Recall by High and Low Imagers

of Stimuli Presented as Words and as Pictures

Francis J. Di Vesta

In an investigation by Jenkins (1963) as cited and described by Stewart (1965) _Ss_ were presented a series of pictures and words. Then, on a subsequent task, the _Ss_ were presented either pictures, words, words _associated_ with pictures seen on the first presentation, and pictures associated with words seen on the first presentation. The _S_'s task was to indicate which of the items in the second series he had seen on the first series. The results demonstrated that pictures were easier to recognize than words, or stated conversely, more errors were made in recognizing words than were made in recognizing pictures. In addition, pictures were mistaken for words less often than words were mistaken for pictures seen before. These findings suggest the greater generalization of words over pictures. Indeed, the authors labeled the tendency to make more errors with words than with pictures as a case of response generalization.

In a follow-up study, Stewart (1965) modified the procedures used by Jenkins, et al. and extended their study by investigating the differences in performance on the recognition task between high and low imagers. The individual differences were defined by scores on the Spatial Relations sub-test of the Differential Aptitude Test Battery (Bennett, Seashore, & Wesman, 1963) and by the Space Thinking (Flags) Test (Thurstone and

PRECEDING PAGE BLANK

Jeffrey, 1959). Stewart summarized her results as follows:

" . . . The pictures were recognized with significantly fewer errors than words. Presenting the items as pictures benefited the high imagers to a greater extent than it did the low imagers; at the same time, both groups were aided significantly [by the pictorial representations]. There was <u>some</u> [emphasis ours] evidence that high imagers were more likely to code a word as a picture than were the low imagers; and vice versa, the low imagers were more likely than were the high imagers to remember or code a picture as a word" (Stewart, 1965, p. 74).

The importance of these findings is in the suggestion that learning materials tend to be coded in the same form they were received by the <u>S</u>. In addition, recognition based on percepts appears to be easier than that based on symbols. Furthermore, there appears to be a tendency for high and low imagers to code materials in different ways, thereby also affecting the retrieval of information. Nevertheless, the theoretical implications of Stewart's results for retrieval are far from clear.

Because of its implications for understanding the processing of incoming information as well as for understanding the interaction between aptitude and individual differences, Stewart's study was replicated as well as extended in the present study. The first phase of the <u>S</u>'s participation was identical to that required of <u>S</u>s in Stewart's study. That is, following an initial presentation of the word and picture stimuli, the materials were presented again, in varying relationships to those presented originally, and the <u>S</u>'s task was to identify those he could recognize as having been presented before. In the second task, which immediately followed the recognition series, the materials on the first task were again presented after which the <u>S</u> was required to recall as many of the items presented in any order he chose; that is the free recall procedure was used. As a replication of Stewart's study the recognition phase of the present study was intended to provide another test of the hypothesis that the coding basis is the form in which the stimuli are presented (i.e., most of the correct responses were expected to be in the form they were presented initially). As an extension of Stewart's study, the present study permitted an examination of the hypothesis that material is retrieved according to the dominant basis of encoding as it is influenced by whichever strategy is reflected in his imagery scores. Thus, it was expected that pictorial material would be encoded more easily than verbal material by high imagers. While the same effects were hypothesized for low imagers (because evidence from earlier studies suggest that pictorial stimuli are learned and recalled more easily than verbal stimuli) it was expected that the difference would not be as great as it would be for high imagers. Furthermore, it was expected that the retrieval preferences of high imagers would be reflected in clustering during recall of pictorial materials to a greater extent than of verbal materials.

Method

Design

The Ss were administered a list of 50 words randomly assorted with 50 pictures representing common objects (Presentation Trial: I). They were then presented another list (Recognition Trial) consisting of 25 words seen before (WW), 25 pictures seen before (PP), 25 words seen as pictures (PW), 25 pictures seen as words (WP), 25 pictures never seen, either as pictures or words, on the presentation trial (NP), and 25 words never seen, either as pictures or words, on the presentation trial (NW). The Ss task was to indicate which items they had seen before, as pictures or as words, and which items had not appeared before. They were then readministered the first list (Presentation Trial: II) after which they were given a 5-min. recall period in which they were to record as many items as possible via free-recall, from the Presentation list. The primary, though not exclusive, analyses were made of PP and WW items correctly recalled in the Recognition Trial and of PPP and WWW items correctly recalled in the Recall Trial. Sex of subject and levels of imagery were included as variables in some of the analyses.

Subjects

One hundred and four female and 80 male educational psychology undergraduates served as Ss in the experiment. From this group the data for groups (n = 25) of men and women high- and low-imagers were selected for analysis. All Ss were informed that the present study was one of several related experiments and that attendance at all sessions was mandatory if credit toward the course grade was to be received.

Measures of Individual Differences

A battery of tests designed to measure imagery ability, verbal ability, and automatization was administered to the entire group of Ss. The tests for measuring imagery consisted of the Flags Test (Thurstone & Jeffrey, 1959), the Spatial Relations Test (Bennett, et al., 1963), and the Gottschaldt Figures Test as described by Thurstone (1944). Verbal ability was measured by a vocabulary test (designed especially for this study by selecting items from existing intelligence tests) and by the College Entrance Examination Board Scholastic Aptitude Test: Verbal

(1962-1963). Automatization was measured by the Stroop Color-Word Test as described by Thurstone (1944) and by an automatization test described by Broverman, Klaiber, Kobayashi, and Vogel (1968). The socres on these tests were standardized for the groups of males and females separately and the resulting T scores for tests comprising a given factor were summed and averaged. The data for the 25 males and 25 females with high scores and the 25 males and 25 females with low socres on these factors were employed in analyses involving aptitude by treatment interactions. The average raw scores for these groups are presented in Table 1. More detailed information on these tests and relationships with other tests administered in the same battery are described in a report by Di Vesta, Ingersoll, and Sunshine (1971).

The analyses presented in this paper are based only on high and low imagers. Other analyses based on high and low automatizers and high and low verbal ability had also been made, In addition, analyses had been made of high and low imagers selected according to the procedure described by Stewart (1965) using only the Flags and Spatial Relations Tests. However, these latter analyses either yielded no effects of consequence or added no further information to the results of the analyses based on the factor scores for imagery. Accordingly, the results of those analyses are not reported here.

<u>Materials</u>

A pool of 300 words from Stewart's (1965) study was used as a basis for preparing stimulus materials. All words were common concrete nouns. Each word had a picture as its counterpart. There were three 100 item lists prepared for Presentation Trial: I. Each list consisted of 50

Table 1

Means of Raw Scores For Men and Women on Tests Comprising Imagery, Verbal, and Automatization Factors in High and Low Groups for Each Factor[a]

Tests	Women		Men	
	Low	High	Low	High
<u>Imagery Factor</u>[b]				
Flags	71.17	121.35	92.64	123.40
Spatial Relations	49.33	88.30	54.00	92.52
Gottschaldt (Total)	27.87	44.35	27.16	46.28
Vocabulary	20.08	18.52	15.88	19.44
SAT (Verbal)	552.23	517.75	478.81	509.45
<u>Verbal Factor</u>				
Vocabulary	15.28	22.16	14.12	21.48
Reading	13.04	19.60	14.24	19.68
SAT (Verbal)	441.64	581.60	441.60	557.24
<u>Automatization</u>[b]				
Stroop (Color-word score)	77.71	114.00	85.76	129.92
Automatization	46.33	62.76	48.64	62.48
Vocabulary	20.86	18.00	18.76	17.16
SAT (Verbal)	572.83	492.64	515.50	500.61

[a] N = 25 in each group.

[b] Vocabulary and SAT (Verbal) scores are presented for these groups to indicate any relationships with the factor tests. Only the differences between high and low automatizers (women) on SAT Verbal scores were significant ($p < .01$).

picture and 50 word stimuli compiled by the random selection of items without replacement from the total pool of words or pictures. Within each list for the Presentation Trial no picture was represented by a word duplicating it. This procedure resulted in three unique lists.

Each of the three lists for the Recognition Trial was constructed to correspond with one of the three series for the Presentation Trial: I. A given list consisted of 150 stimuli, 75 of which were pictures and 75 were words. The items were selected so that half (n = 25) of the word stimuli used in the Presentation Trial: I were maintained as words for the Recognition Trial while the other half (n = 25) were represented pictorially. Similarly, half (n = 25) of the original pictures were maintained as pictures for the Recognition Trial while the other half (n = 25) were represented as words. The remaining 50 items were divided equally between words and pictures none of which had either verbal or pictorial counterparts in the rest of the list; nor had these 50 items been seen, in either form, on Presentation Trial: I. Thus, each Recognition list consisted of 50 stimuli (25 words, 25 pictures) in the form originally seen on the Presentation Trial: I, 50 stimuli in converted (associated) form, and 50 stimuli previously unseen (new), 25 of which were pictures and 25 of which were words. The stimuli were ordered randomly, via reference to a table of random digits, within the lists.

Procedure

The task was administered to Ss in groups varying in number from 2 - 8. However, each S worked individually and at separate locations (every other seat) in a room approximately 3 x 5 meters in size. The Ss were seated opposite a large screen and were informed that they would be participating in an experiment on memory. They were told that a series of slides, showing pictures and words, would be projected onto the screen and that their task would be to study the slides and attempt to remember as many items as possible. Following these instructions, one of the three Presentation Trial lists, selected at random, was projected at a rate of 3-sec. per slide.

Immediately after Presentation Trial: I the Ss were provided with an answer sheet on which were spaces for indicating whether a slide to be presented on the Recognition Trial had been seen on Presentation Trial: I and, if so, whether it had been seen as a picture or a word. Each slide on the Recognition trial was presented at a rate of 8-sec. per slide. The S was informed that his task on the Recognition Trial was to decide whether the item on the slide a) was completely new, i.e., it was not seen in the presentation series either as a picture or a word; b) was seen previously as a word or; c) was seen previously as a picture, and to mark his (or her) answers accordingly on the answer sheet. The Ss were not told the proportion of slides represented in each category. Although guessing was encouraged where the S was uncertain, E also expressed the need for rapid and accurate decisions. In order to aid the S in keeping his place, the slides for the Recognition Trial were ordered numerically and E called out the number of each slide as it appeared on the screen.

After the Recognition Trial was completed, Presentation Trial: II, consisting of the same series as Presentation Trial: I, was administered. The Ss were instructed to study each item carefully and to remember as many items as possible. At the completion of the slide presentation, blank sheets of paper were distributed for the Free-Recall of all items (whether picture or word) that the S could remember from Presentation Trial: II and in any order he chose. The Ss were not required in the Recall Phase to indicate whether the item recalled had been presented as a picture or a word during Presentation Trial: II.

Results

The primary dependent variables, based on responses to the Recognition Trial, were: a) the number of items recalled and labeled correctly; and b) the number of items said to have occurred as words or pictures on Presentation Trial: I when in fact they had not been presented at all. The dependent variable based on responses to the Free-Recall task was the number of items, presented as words or pictures on Presentations I and II, that were recalled correctly. Pairs of pictures or of words, scored as adjacencies, were used as measures of clustering in further analyses.

All analyses of data were made by a mixed analysis of variance design. However, the variables used differed from one analysis to another. Typically, the between-subjects variables were the randomizations of the items (Item Series), Imagery Ability, and Sex of subject. The within-subjects variables for analyses of responses made on the Recognition Trial consisted of the Form of the Stimuli on the Presentation Trial: I and the Form of Stimuli on the Recognition Trial. In addition, the analyses of responses on the Free-Recall Task included the form of the stimuli on the Presentation Trial: II as another within-subjects variable. Inasmuch as it was virtually impossible to interpret significant main effects or interactions involving the Item Series, their occurrence will be noted in the reporting of results without further comment.

Recognition

The analyses of the number of items correctly recognized on the Recognition Trial by female High and Low Imagers are summarized in Table 2. The parallel analysis for the male High and Low Imagers are summarized in Table 3. In both tables comparisons have been made between combinations of words and pictures shown on Presentation Trial: I and the Recognition Trial. Thus, words on both trials (WW) have been compared with pictures on both trials (PP); new words (NW) of objects (i.e., a new set of words was presented on the Recognition Trial the items of which had not appeared either as words or as pictures on the Presentation I trial) were compared with new pictures (NP) on the Recognition Trial; and words on Presentation Trial: I followed by pictures on the Recognition Trial (WP) were compared with pictures followed by words (PW).

In these analyses the main effects of Item Series was significant only for the comparison of responses by male Ss to the WW and PP stimuli (F [2,54] = 3.84, $p < .05$). However, Item Series interacted with Imagery Level in the comparison of responses by female Ss to the WW and PP stimuli (F [2,54] = 5.04, $p < .05$) with Form of Stimuli in the comparison of responses by male Ss to the NW and NP stimuli (F [2,54] = 4.90, $p < .05$).

The most important and clear cut effects to be noted in these results, however, are in the comparisons between Form of Stimuli. In all analyses (see Table 2) the differences were highly significant ($p < .001$). As shown in Table 4, for both men and women Ss, performance on PP stimulus combinations was more accurate than on the WW combinations; performance

Table 2

Summary of Analysis of Variance of the Number of Items Correctly Recognized On the Recognition Trial by Women High and Low Imagers

		Comparison					
		WW vs. PP		NW vs. NP		WP vs. PW	
Between Subjects	df	MS	F	MS	F	MS	F
Item Series (A)	2	20.03	1.00	.41	.02	10.68	0.55
Imagery Level (B)	1	3.68	0.18	2.13	.08	7.01	0.36
A x B	2	100.80	5.04[b]	4.35	.17	32.60	1.69
Error (b)	54	20.02		25.90		19.26	
Within Subjects							
Form of Stimuli (C)	1	516.68	82.36[d]	163.33	19.81[d]	156.41	14.45[d]
A x C	2	12.70	2.03	36.26	4.40[b]	9.16	0.85
B x C	1	6.08	0.97	10.80	1.31	9.08	0.84
A x B x C	2	7.30	1.16	1.08	0.13	9.03	0.83
Error (w)	54	6.27		8.24		10.83	

[a] $p < .10$

[b] $p < .05$

[c] $p < .01$

[d] $p < .001$

Table 3

Summary of Analysis of Variance of the Number of Items Correctly Recognized on the Recognition Trial by Male High and Low Imagers

		Comparison					
		WW vs. PP		NW vs. NP		WP vs. PW	
Between Subjects	df	MS	F	MS	F	MS	F
Item Series (A)	2	99.99	3.84[b]	7.06	0.19	49.03	1.56
Imagery Level (B)	1	2.41	0.09	33.21	0.97	116.03	3.69[a]
A x B	2	12.86	0.49	87.86	2.41	21.73	0.69
Error (b)	54	26.03		36.43		31.43	
Within Subjects							
Form of Stimuli	1	715.41	56.85[d]	69.01	10.09[d]	145.20	15.21[d]
A x C	2	26.66	2.12	35.51	4.90[b]	.40	0.04
B x C	1	1.86	0.15	.01	0.00	3.33	0.35
A x B x C	2	8.18	0.65	7.11	1.04	10.03	1.05
Error (w)	54	12.58		6.84		9.55	

[a] $p < .10$

[b] $p < .05$

[c] $p < .01$

[d] $p < .001$

Table 4

Mean Numbers of Correct Responses on Recognition Trial

Form of Stimuli on Presentation I and Recognition Trials	Sex of Subject		
	Men	Women	Row Means
Word-Word	15.83	17.08	16.46
Word-Picture	15.72	17.23	16.48
Picture-Word	17.92	19.52	18.72
Picture-Picture	20.72	21.23	20.48
None-Word	17.87	18.83	18.35
None-Picture	16.35	16.50	16.43

on the PW order of presentation was clearly better than on the WP order; and new words (NW) were recognized more accurately as not having appeared on the presentation trial than were new pictures (NP). In every case, without exception, the performance of women on these tasks exceeded the accuracy of the performance of men.

Similar analyses of recognition errors are summarized in Table 5. The reader should note that these analyses are not of mere incorrect recognitions which would be only the difference between actual and possible correct responses. Instead, this analysis is of the errors made in labeling a stimulus on the Recognition Task as having been presented as a word or picture or Presentation Trial: I or as not having been seen on Presentation Trial: I. If a picture on the Recognition Trial was said to have been presented as a picture on Presentation Trial: I, when indeed, it had been presented as a word, this was called a word-picture error (WPER). If a word presented on the Recognition Trial was said to have been presented as a word on Presentation Trial: I when it actually had been presented initially as a picture, this was called a picture-word error (PWER). NWER and NPER refer to errors in which a word or picture appearing on the Recognition Trial were incorrectly recognized as having been presented on Presentation Trial: I. Thus, NWER errors refer to the number of times an S said that a word had appeared on Presentation Trial: I as either a word or picture when it had not been so presented; conversely, NPER errors refer to the number of times an S said that a picture had appeared on the Presentation Trial: I as either a word or picture when it had not been so presented.

The analyses summarized in Table 4 yielded significant main effects due only to comparisons between Form of Stimuli. The comparison of WPER

Table 5

Summary of Analyses of Variance of Errors Made on Recognition Trial

Source of Variance		Word-Picture (WP) Errors vs. Picture Word (PW) Errors				New Words (NW) vs. New Picture (NP) Errors			
		Females		Males		Females		Males	
Between Subjects	df	MS	F	MS	F	MS	F	MS	F
Item Series (A)	2	.75	.05	38.43	1.36	.41	.02	7.06	.19
Imagery Level (B)	1	7.50	.50	13.33	.47	2.13	.08	35.21	.97
A x B	2	25.98	1.74	34.53	1.22	4.36	.17	87.85	2.41
Error (b)	54	14.91		28.21		25.91		36.43	
Within Subjects									
Form of Stimuli (C)	1	22.53	5.75[b]	3.33	.36	986.13	62.34[d]	1122.42	64.96[d]
A x C	2	8.66	2.21	38.03	4.10	22.16	1.40	59.16	3.42
B x C	1	1.63	.41	6.53	.70	8.53	.54	2.41	.14
A x B x C	2	.51	.13	4.93	.53	15.41	.97	4.65	.27
Error (w)	54	3.92		9.28		15.82		17.28	

with PWER errors yielded F (1,54) = 5.75, $p < .05$ for data based on responses of female Ss. Though this main effect was not significant for male Ss, there is some indication that the Item Series may have somehow been influential in this regard inasmuch as the interaction of Item Series and Form of Stimuli yielded F (2,54) = 4.10, $p < .05$ for data based on male Ss. The comparison of NWER and NPER were clearly the same for female and male Ss since the analyses yielded F (1,54) = 62.34 and F (1,54) = 64.96, respectively, both of which are highly significant ($p < .001$).

The means for these comparisons are presented in Table 6. There it may be seen that there were relatively more errors caused by mislabeling a word on the Recognition Trial when it had been presented initially as a picture, than in mislabeling a picture on the Recognition Trial when it had been presented as a word on Presentation Trial: I. A major error of the Ss was to indicate that a word presented on the Recognition Trial had been presented on the Presentation Trial: I when it had not been. The data related to these findings support those reported by Stewart.

Free Recall

The analyses of variance of the number of items correctly recalled on the Free-Recall task are summarized in Table 7 for the data based on the responses of female Ss and in Table 8 for data based on the responses of male Ss.

The result of primary interest in these analyses is the effect due to the interaction of Imagery Levels with Form of Stimulus on Presentation Trial: I. This effect yielded F (2,108) = 3.10, $p < .05$ for the data

Table 6

Mean Number of False Recognitions on the Recognition Trial

by Men and Women High and Low Imagers

Presented on Presentation Trial: I and Recognition Trial as:	Described on Recognition Trial as having been presented on Presentation I as:	Low Imagers		High Imagers		Total	
		Men	Women	Men	Women	Men	Women
Word-Picture	Picture	5.17	2.90	4.03	3.17	4.60	3.03
Picture-Word	Word	5.03	3.53	4.83	4.27	4.93	3.90
Not Presented	Word	11.63	9.80	10.27	10.60	10.95	10.20
Not Presented	Picture	5.23	4.60	4.43	4.33	4.83	4.47

Table 7

Summary of Analysis of Variance of the Number of Items Correctly Recalled on the Free-Recall Task by Female High and Low Imagers

Between Subjects	df	MS	F	
Item Series (A)	2	3.85	0.35	
Imagery Level (B)	1	25.60	2.30	
A x B	2	1.61	0.15	
Error (b)	54	11.11		
Within Subjects				
Form of Stimuli on Presentation I (C)	2	3561.20	602.74	$p < .001$
A x C	4	47.97	8.12	$p < .01$
B x C	2	18.31	3.10	$p < .05$
A x B x C	4	1.58	0.27	
Error (w_1)	108	5.91		
Form of Stimuli on Presentation II (D)	1	5.38	1.43	
A x D	2	18.04	4.78	$p < .05$
B x D	1	1.88	0.50	
A x B x D	2	1.64	0.43	
Error (w_2)	54	3.77		
C x D	2	18.43	2.97	$p < .10$
A x C x D	4	24.87	4.01	$p < .01$
B x C x D	2	4.80	0.78	
A x B x C x D	4	1.90	0.31	
Error (w_3)	108	6.20		

Table 8

Summary of Analysis of Variance of the Number of Items Correctly Recalled on the Free-Recall Task by Male High and Low Imagers

Between Subjects	df	MS	F	
Item Series (A)	2	14.17	1.27	
Imagery Level (B)	1	96.10	8.61	$p < .01$
A x B	2	.98	0.09	
Error (b)	54	11.16		
Within Subjects				
Form of Stimuli on Presentation I (C)	2	2786.32	391.47	$p < .001$
A x C	4	31.77	4.46	$p < .01$
B x C	2	47.11	6.66	$p < .01$
A x B x C	4	8.77	1.16	
Error (w_1)	108	7.12		
Form of Stimuli on Presentation II (D)	1	1.11	0.28	
A x D	2	8.30	2.08	
B x D	1	2.18	0.55	
A x B x D	2	7.15	1.79	
Error (w_2)	54	3.99		
C x D	2	8.05	1.75	
A x C x D	4	3.82	8.31	
B x C x D	2	.29	0.06	
A x B x C x D	4	.92	0.20	
Error (w_3)	108	4.60		

based on the responses of women Ss and F (2,108) = 6.66, $p < .01$ for the data based on the responses of men. In addition, the main effect associated with Imagery Levels for men yielded F (1, 54) = 8.61, $p < .01$.

In Table 9 it can be seen that the women recalled more items than men after Presentation Trial: II. However, of more interest is the finding that recall a) is directly related to the form of stimuli on the two presentations and b) is affected by the form of the stimuli on the recognition trial. Thus, W-W presentation (i.e., word on Presentation Trial: I and Presentation Trial: II, disregarding the form on the Recognition Trial) results in poorer performance than a P-P presentation as can be seen in Table 10. However, either W-W or P-P presentations, when interpolated with the other form on the Recognition Trial (thus, either WPW or PWP) results in better free-recall than when the same form is used in the Recognition Trial (that is, either WWW or PPP).

The data presented in Table 11 are related to the interaction of Imagery Levels with the Form of Stimuli on the Presentation Trials. There it can be seen that High Imagers (both men and women) recall more items, in general, than do Low Imagers. However, in both groups (that is, men and women) High Imagers recall more items depicted pictorially than do the Low Imagers. These data, however, must be interpreted cautiously. The direct relationship between the difference in word and picture recall and the total recall scores (or either word or picture recall scores singly) suggests that Imagery Levels may be reflecting a general intellectual factor. Consequently, the data may simply imply that the person with the higher ability is able to profit more from the optimal treatment which, presumably, is the pictorial presentation of stimuli.

Table 9

Mean Numbers of Correct Responses Recalled on Free-Recall Trial

As a Function of the Effect Due to Form of Stimuli on Presentation Trials

Form of Stimuli on Presentation Trials	Sex of Subject Men	Women
Words	8.78	9.71
Pictures	10.08	11.24
New*	1.15	1.13

* These are intrusions from the Recognition Trial. These words or pictures were present on the Recognition Trial only and were not present on either Presentations I and II.

Table 10

Mean Number of Correct Responses on Free Recall Trial

Form of Stimuli on Presentation Trial: I, Recognition Trial, and Presentation Trial: II	Sex of Subject	
	Men	Women
Word-Word-Word	8.55	9.20
Word-Picture-Word	9.02	10.21
Picture-Picture-Picture	9.83	10.97
Picture-Word-Picture	10.32	11.51
None-Word-None	.93	1.00
None-Picture-None	1.33	1.27

Table 11

Mean Number of Items Correctly Recalled by Men and Women High and Low Imagers On Free-Recall as a Function of Form of Stimuli on Presentation Trials

Form of Stimuli	High Imagers			Low Imagers		
on Presentation Trials	Men	Women	Total	Men	Women	Total
Words	9.63	10.08	9.86	7.93	9.33	8.63
Pictures	10.98	11.83	11.40	9.17	10.65	9.91
Not Presented	0.95	0.97	0.96	1.37	1.30	1.33
Diff (p-w)	1.35	1.75	1.55	1.24	1.32	1.28

A final analysis was made of the proportion of adjacent pictures or words in the free-recall task. These were computed simply by dividing the number of observed pairs of pictures or words by the number of possible pairs of pictures or words respectively. None of the main effects or interactions involving these data were significant ($p > .05$). However, since the trends for the interaction between Imagery Level and Form of Adjacent Pairs were very similar to those reported for similar data by Stewart (1965) they are presented in Table 12. The reader should note that this interaction is based on data for all Ss and not just extreme groups used in previous analysis. The interaction yielded $F(1,116) = 2.63$, $p < .20 > .10$. If this finding can be considered a reliable one since it replicates Stewart's findings, it implies that Low Imagers tend to organize (retrieve) materials presented in verbal form more efficiently than materials presented in pictorial form. On the other hand, the data for High Imagers imply greater organization of pictorially presented materials than of verbally presented material.

Discussion

This experiment has clearly replicated the earlier findings by Stewart (1965) with regard to the recall of materials presented in picture and word forms. In addition, it provides clear evidence for the differential performance of men and women on the kinds of tasks that were used. On the other hand, though there was some slight evidence that Ss classified as high and low Imagers perform differently on materials presented in picture or word forms the data can only be considered as suggestive.

Table 12

Proportions (Observed/Possible) of Adjacent Pictures or Words in the Free-Recall of High and Low Imagers

	Proportion of	
Imagery Level	Word Pairs	Picture Pairs
Low	.14	.12
High	.12	.18

Perhaps the most important implication of the present study is to be obtained from an integration of the findings from the recognition and free-recall phases which suggest the relative roles of imaging and labeling in storage and retrieval processes in memory. Since pictures were more frequently recalled than words there is the implication that words, as highly generalized symbols, refer to idealized attributes or qualities and consequently create more interference in recall than do pictures. Speculation regarding the specificity of the referent for pictures has been made in the introduction to this paper. Since pictures are relatively more distinctive than words it can be assumed they will suffer less from interference during the Recognition Trial.

The findings also clearly imply that the picture-to-word order results in more correct identifications than the word-to-picture order. This effect may be the result of factors associated with developmental processes, i.e., Ss in the college culture have had more experience in providing verbal labels for pictures than in providing images for words. Nevertheless, for whatever reason, the finding that providing a label for a picture does facilitate recognition more than providing a picture for the word is a clear and reliable one. It would appear, by way of explanation, that the verbal response was included in responses to the picture but the percept or imaginal (picture) response was less likely to be included in responses to the verbal label for the referent. This explenation indicates that redundancy on the Presentation, Recognition, and Recall Trials (e.g., word-word or picture-picture) is not as efficient for recall as is bi-modal presentetion. These explenations elso apply to the findings that new words are more easily recognized than new pictures. Thus, because there are fewer words end more pictures incorporated into the S's memory during the first presentation, there will be less interference to new words than to new pictures.

In general, the findings from the present study imply that materials are received first as images and then verbally labeled for storage. If this assumption is correct, it would also be expected that the picture-word order of presentation would yield more correct responses in free recall than the word-picture order of presentation, because the latter requires S to perform an additional operation of reversing the order. The picture-word order should also be superior to the picture-picture order which does not provide for labeling thereby hindering effective storage, or the word-word order which does not provide a percept to make the label easily discriminable (less generalized). These assumptions were provided substantial support in this study. The findings are especially interesting since the PPP was not the most efficient order of presentation as would have been suggested if only the Ss' performance on the recognition trial had been investigated.

The aforementioned findings and assumptions imply a dual process in retrieval of information. On the one hand, recognition depends on the distinctiveness of percepts, in which case the pictures are more easily identified than words. On the other hand, retrieval depends on encoding processes or the strategy by which materials are stored, in which case the picture which has been labeled by a word is more easily retrieved than words followed by words (which are subject to interference); or than words followed by pictures because pictures presumably add little to the distinctiveness of a generalized symbol; or than pictures followed by pictures because images or percepts are retained only for brief periods of time unless they are labeled.

The superiority of the female Ss recall over the males recall is undoubtedly due to factors associated with developmental processes. It is too early to say what the nature of the factors that account for this differential performance might be.

None of the results of the present study provided support for Stewart's (1965) findings which she summarizes as follows:

"High imagers were superior to low imagers in picture recognition but inferior in word recognition. Though the picture recognition was significantly better for both types of imagers than was the word recognition, the high imagers benefitted to a greater extent."

Part of the reason for lack of supporting evidence may have been that the Ss in the present study achieved higher mean recognition and recall scores than those in Stewart's study. Although the interactions associated with Imagery Ability on the Free-Recall task implied that high Imagers benefited more by pictures than by words, the alternative explanation remained that such differences might be attributable to other ability factors associated with Imagery. Furthermore, while the finding that low Imagers tend to organize by words in free-recall and high Imagers tend to organize by pictures supports the tendencies of the two groups suggested in Stewart's study, the differences were not significant.

Thus, we are led to conclude that the present results point to a clear superiority of pictures over words for presentation of learning material related to a given referent if that material is to be recognized easily on later occasions. Furthermore, the superiority of the picture to word over the word to picture order suggests that percepts are provided verbal labels for storage and that such labels facilitate later retrieval as measured by the number of items recalled on a free recall task. If these treatment variables interact with Imagery levels to affect recognition or recall, the effect must be considered to be a fragile one or of limited generalizability at best.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential aptitude tests. (Grades 8-13 and adults.) New York: Psychological Corporation, 1963.

Broverman, D. M., Klaiber, E. L., Kobayashi, Y., & Vogel, W. Roles of activation and inhibition in sex differences in cognitive abilities. Psychological Review, 1968, 75, 23-50.

College Entrance Examination Board Scholastic Aptitude Test. Princeton, N. J.: Educational Testing Service, 1962-1963.

Di Vesta, F. J., Ingersoll, G., & Sunshine, P. A factor analysis of imagery tests. Journal of Verbal Learning and Verbal Behavior, 1971, in press.

Jenkins, J., Carlson, Tweedy and Doren. Response generalization. University of Minnesota, Unpublished manuscript, presented at the Second California Conference on Verbal Learning and Verbal Behavior, 1963.

Lockhart, R. S. Retrieval asymmetry in the recall of adjectives and nouns. Journal of Experimental Psychology, 1969, 79, 12-17.

Paivio, A. Learning of adjective-noun paired-associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370-379.

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, The University of Toronto, 1965.

Thurstone, L. L. A factorial study of perception. Chicago: University of Chicago Press, 1944.

Thurstone, L. L., & Jeffrey, T. G. Space thinking (Flags). Chicago; Education-Industry Service, 1959.

Yuille, J. C., Paivio, A., & Lambert, W. E. Noun and adjective imagery and order in paired-associate learning by French and English subjects. Canadian Journal of Psychology, 1969, 23, 259-466.

Verbal and Imaginal Processing in Learning and Transfer by High and Low Imagers

Francis J. Di Vesta

Both Kuhlman (1960) and Stewart (1965) found that high imagers learned a paired-associate list most easily when the stimuli were pictures than when they were words. Furthermore, Stewart (1965) found that high imagers learned tasks in which pictorial stimuli were used more easily than did low imagers while low imagers learned lists in which verbal stimuli were used in fewer trials than did high imagers. These results imply that investigations of aptitude by treatment interactions (ATI) may be a useful method for investigating strategies employed by _Ss_ during learning and recall.

The present study was first of all an attempt to replicate, in part, one of Stewart's experiments. Such replication appears to be especially justifiable in view of Cronbach and Snow's comments, as follows:

> . . . Progress toward the goal of identifying and understanding ATI has been slight. We have not examined every pertinent study but our survey has probed deeply enough to give us confidence that a truly exhaustive sample would not change the general picture as of this moment. _There are no solidly established ATI relations_ even on a laboratory scale and no real sign of any hypothesis ready for application and development. There are intriguing findings here and there, _none of which has been pursued through a sufficient series of replication_, validity

generalization, and enhancement studies to make it impressive (1969, p. 193; italics ours).

Secondly, the major purpose of the present study was to investigate the effects of learning with pictorial and verbal stimuli on performance of a transfer task. It can be reasoned that pictorial representations tend to be more specific than a verbal stimulus for the same referent. Thus, while the pictorial stimulus, <u>bird</u>, will, under most circumstances, elicit the verbal mediator "bird" it will also be fairly specific in the sense that it cannot be a "template" for the abstraction of "all birds" unless it achieves symbolic status, as it might in abstract or expressionistic painting, at which point its equivalence to a verbal symbol can be assumed. In any diagrammatic representation there will be some restriction on what is perceived even though it is nothing more than the restriction that a class of "large birds" or of "small birds" is represented. On the other hand, the verbal stimulus "bird," or any other similar symbol for that matter, is more nearly representative of a highly generalized "template." In the sense that it represents a larger class of all experiences the <u>S</u> has had with birds, a verbal symbol should provide a broader base for transfer than the pictorial stimulus.

Based on the aforegoing rationale the present study extended Stewart's investigations through an experiment designed to examine the extent to which mediation processes differentially involved imaginal or verbal transformation of an experience. More specifically, the intent was to investigate whether one form of mediation takes <u>precedence</u> over another. It was hypothesized that if imagery was a dominant processing mechanism for storing and using concrete materials then transfer to pictorial representations (geometric representations or their verbal equivalents) should be facilitated when compared with transfer to verbal representations (particularly verbal subordinates). If verbal mediation was more advantageous than imaginal processing then, it was hypothesized, transfer to verbal representations would be more efficient than to pictorial representations.

It was also possible in this study to investigate a sequencing hypothesis. Thus, it was hypothesized that both imaginal and verbal mediation might be employed in processing and, if so, one possibly precedes the other. Accordingly, if imaginal precedes verbal mediation then the pictorial to verbal order of presentation was hypothesized to be more efficient than the verbal to pictorial order, and vice versa (Lockhart, 1969; Paivio, 1963; Yuille, Paivio, & Lambert, 1969). The hypotheses related to imagery, as an individual difference variable, extend the above hypotheses by implying that high imagers would be especially benefited by any treatment favoring imaginal processing while low imagers would be benefited by any treatment favoring verbal processing.

Method

General Design

Nouns which could also be graphically presented by basic geometric shapes served as stimuli during original learning. The *S* learned to associate either the verbal or pictorial representations of each of these nouns with a number from 2 through 9. The form of the stimuli (pictures or words) served as a between-subjects dimension. The *S*s of each sex, which served as the other between-subjects dimension, were assigned in equal numbers to each experimental treatment.

The transfer task consisted of four lists in which the stimuli represented one of four relationships to the stimuli in the original learning task list. Thus, the lists for the transfer task were comprised of verbal subordinates, geometric representations, verbal equivalents of geometric representations, and identical representations but opposite modes of presentation of the object or word presented during initial learning. Each _S_ received all four types of transfer stimuli which comprised a within-subjects dimension. Trials to criterion during original learning and number of correct responses to stimuli in each of the four transfer conditions served as the dependent measures. In addition to the mixed analyses of variance of these data, the effect of each of four aptitude variables on the number of correct responses was analyzed.

Stimulus Materials

Word (W) and picture (P) forms of eight generic nouns were paired with the digits 2 through 9 for the lists used in the original learning (OL) task. The transfer stimuli were either words (W) or pictures (P) in four relationships to the stimuli in the OL task. Thus, the transfer lists were: Condition SN, subordinates of the generic nouns presented during OL, e.g., the word or picture, dime, was used in the transfer task as a subordinate of the word coin used in the OL task; b) Condition GN, geometric representations or outline drawings of the generic noun presented during OL, e.g., a circle was used as a representational drawing of a coin; c) Condition GL, labels or word equivalents for the geometric figures used in Condition GN, e.g., the word "circle" rather than the figure might be presented in Condition GL; and Condition D, in which the representations in the transfer task were identical to those in the OL task but were presented in the opposite mode during OL, e.g., if a picture of a coin was presented during OL, the word "coin" would serve as a transfer stimulus, and vice versa. The stimuli used in all conditions are presented in Table 1.

Subjects

Undergraduate students at The Pennsylvania State University who were enrolled in an introductory course in educational psychology served as Ss. Although they were volunteers, the Ss received credit toward their final grade for participating in the experiment. All Ss had been previously administered a battery of tests designed to measure imagery and verbal ability. Assignment to learning conditions was randomized within sex of Ss. In all, 198 Ss, of which 94 were male, participated in this experiment.

Measures of Individual Differences

Measures of individual differences were obtained for Imagery-Ability, Verbal ability and Automatization. The Flags test (Thurstone & Jeffrey, 1959), Spatial Relations Test of the Differential Aptitude Test (Bennett, et al., 1963), and the Gottschaldt Figures Test as described by Thurstone (1944) were employed to measures Imagery-ability. The Stroop Color-Word test as described by Thurstone (1944) and the Automatization test (Broverman, Klaiber, Kobayaski, and Vogel (1968) were used to measure Automatization. A vocabulary test, reading test (both of which were locally constructed by compiling items from existing tests) and the College Examination Board's Scholastic Aptitude Test (1962-1963). The raw scores for tests associated with each factor were standardized for males and females separately. The resulting T-scores were summed and averaged to obtain an overall factor score. High- and

Table 1

Original and Transfer Stimuli for the

Learning and Transfer Conditions

	Original Learning	Transfer*	
Response	Generic Noun	Subordinate Noun	Word for Geometric Representation
2	Coin	Penny	Circle
3	Horn	Bugle	Cone
4	Cheese	Swiss	Wedge
5	Flower	Rose	Star
6	Snake	Python	Coil
7	Gem	Ruby	Hexagon
8	Tree	Spruce	Triangle
9	Drum	Snare	Cylinder

* Note -- The figures for the generic nouns were pictures of the objects listed under original learning. The figures for the geometric representations were pictures of the forms shown in the last column.

low-scorers were selected from the extremes of the resulting distributions of factor scores for men and women. High scorers had average T-scores greater than 55 and low scorers had average T-scores lower than 45.

Procedure

Upon arrival at the laboratory the S was randomly assigned to one or the other of the OL conditions. Instructions were provided as to the specific nature of the task, i.e., the S was either told that he would be presented word-number pairs or that he would be presented picture-number pairs. The study-test procedure with standard instructions was used.

During the study trial, stimulus pairs were rear-projected onto a translucent screen by a carousel projector at a 2-sec. rate. During the recall interval only the stimulus member was presented at a 2-sec. rate. The S's task in the recall phase was to respond with the number previously associated with it. Study-recall trials were administered until the S had identified all but one of the paired-digits correctly.

The transfer task was administered following a rest period of 2-min. The S was instructed that he would be presented 32 stimuli most of which he had not seen before but all of which had some relationship to the words learned during the OL task. He was instructed to respond with the same digits employed in OL and was told to base his response on possible relationships to the list he had just learned. Only one presentation of the transfer list was administered.

Results

Original Learning

The number of trials to criterion on the initial learning task were analyzed by a mixed analysis of variance. This analysis yielded F (1,194) = 9.68, p < .01 for the effect due to the kind of stimuli employed. The list based on words took more trials to learn ($\bar{X}$ = 3.50) than did the list based on pictures ($\bar{X}$ = 2.64). Thus, while the referent in each instance was the same for pictures and for words, pictorial depiction of stimuli clearly resulted in more rapid learning than symbolic presentation. These results suggest differential processing of stimuli presented by the two methods. The effects related to Sex of Subject or its interaction with the task variable were not significant (p > .05).

Transfer Performance

The number of correct responses for the transfer task were analyzed via a mixed analysis of variance with Sex of Subject and Mode of Presentation (W or P) during OL as the between-subjects variables and Kind of Stimulus (i.e , SN, GN, GL, and O) as the within-subjects variable. This analysis yielded F (1,94) = 12.63, p < .001, for the effect due to the Mode of Presentation during OL; F (1,194) = 3.82, p = .05 for the effect associated with Sex of Subject; F (3,582) = 358.07, p < .001 for the effect due to the Kind of Stimulus employed in the transfer task; and F (3,582) = 7.20, p < .001 for the interaction between Mode of Presentation during Original Learning and Kind of Stimulus employed on the transfer task. A summary of this analysis is presented in Table 2.

Table 2

Summary of Analysis of Variance

Of the Number of Correct Responses for All Subjects

Between Subjects	df	MS	F	
Sex (A)	1	27.01	3.82	$p = .05$
Presentation Mode:OL(B)	1	89.17	12.63	$p < .001$
A x B	1	2.25	0.32	
Error (b)	194	7.06		
Within Subjects				
Kind of Stimuli:Transfer (C)	3	478.04	358.07	$p < .001$
A x C	3	2.46	1.84	
B x C	3	9.62	7.20	$p < .001$
A x B x C	3	1.31	0.98	
Error (w)	582	1.34		

The women made more correct responses ($\overline{X} = 4.77$) on the transfer task than did the men ($\overline{X} = 4.39$). The means represented in the interaction between Mode of Presentation during OL and transfer are summarized in Table 3. The primary implication of the data in this table is that pictorial presentation of stimuli during original learning results in greater transfer than does verbal (symbolic) presentation for all transfer conditions; the difference, however, is least when the opposite mode or subordinate representation is employed in the transfer task. It was particularly disadvantageous for Ss to learn words on the first task and then transfer to a geometric representation whether that representation was in symbolic or pictorial form.

Individual Differences

The investigation of effects associated with individual differences was made by extending the design to include the high and low imagers as a third between-subjects factor. There were 10 Ss in each cell of this design.

This analysis yielded $F(1,72) = 13.20$, $p < .01$, for the effect associated with Imagery. The mean score for high-Imagers was higher ($\overline{X} = 4.90$) than that for low Imagers ($\overline{X} = 4.15$) on the transfer task. The hypothesized interaction between individual differences in visualization and treatments was not supported. None of the interactions was significant ($p > .05$).

The same analysis as that described immediately above was conducted by replacing the groups of Ss differing on Imagery scores with groups of Ss differing on Automatization scores. There were also 10 Ss in each group for this analysis, which is summarized in Table 4. The unique outcome of this analysis, compared to the earlier one, was that the third-order interaction was significant (F [3,216] $= 3.13$, $p < .05$).

Table 3

Mean Number of Correct Responses on the Transfer Task

As Related to the Original Learning Task

Transfer Task Conditions	Presentation Mode: OL Words	Pictures	Difference
Opposite Modality	6.15	6.34	.19
Subordinate Representation (Words and Pictures Combined)	5.34	5.76	.42
Verbal Equivalent of Geometric Figure	2.55	3.51	.96
Geometric Figure Representation	2.90	4.03	1.13

Table 4

Summary Analysis of Variance Based on the Mean Number Of Correct Responses for High and Low Automatizers

Between Groups	MS	df	F	
Sex of Subject (A)	0.53	1	0.08	
Presentation Mode: OL(B)	52.00	1	8.04	< .01
Automatization (C)	22.58	1	3.49	< .10
A x B	1.65	1	0.26	
A x C	.01	1	0.00	
B x C	9.45	1	1.46	
A x B x C	.38	1	0.59	
Error (b)	.65	72		
Within				
Kind of Stimuli: Transfer (D)	215.14	3	180.24	< .01
A x D	1.99	3	1.66	
B x D	8.78	3	7.35	< .01
C x D	.57	3	0.48	
A x B x D	2.54	3	2.13	< .10
A x C x D	.38	3	0.31	
B x C x D	.84	3	0.70	
A x B x C x D	3.73	3	3.13	< .05
Error (w)	1.19	216		

The means for the groups represented in this interaction are summarized in Table 5. As in the previous analyses, all groups were found to function more effectively when pictures were employed in OL than when words were used. However, the main difference among groups to be identified in this table is the generally inferior performance of the women who were high Automatizers (i.e., those who take longer to perform the automatization tasks) relative to the performance of women who were low Automatizers (i.e., those who perform the automatization tasks in a shorter period of time) on all transfer tasks following original learning with word stimuli. The women low Automatizers who learned words on the original learning task averaged 1.10 more correct answers on the Transfer Task than did the high Automatizers. The average difference in performance on the transfer task between the same groups when pictures were used as stimuli during original learning was .03. Similar comparisons for males yielded averages of .65 (with words as stimuli in OL) and .47 (with pictures as stimuli) correct responses. If automatization can be considered as a measure of one kind of Imagery these results imply that the encoding by high-Imagers of verbal stimuli is more detrimental to their performance on new tasks than the encoding of pictorial stimuli. This difference holds for both women and men but less so for men.

Comparable analyses based on two other individual difference grouping, verbal and anxiety, were also made. The analysis of verbal groups yielded $\underline{F}$ (1,72) = 5.96, $\underline{p}$ < .05 for differences in performance on the transfer task, associated with verbal ability. As would be expected, the group with the higher verbal ability made more correct responses ($\bar{X}$ = 4.81) than did the groups with the lower verbal ability

Table 5

Mean Number of Correct Responses on the Transfer Task as Functions of Presentation Mode During Original Learning, Kind of Presentation During Transfer, Sex of Subject and Score on the Automatization Factor

Presentation Mode: OL	Transfer Task	Females High Automatizers	Females Low Automatizers	Difference	Males High Automatizers	Males Low Automatizers	Difference
Words	Opposite Modality	6.00	7.00	(1.00)	6.00	6.50	(0.50)
	Subordinate	5.40	6.50	(1.10)	4.60	5.10	(0.50)
	Geometric Figure	2.30	3.00	(0.70)	2.20	3.50	(1.30)
	Verbal Representation	1.50	3.10	(1.60)	2.40	2.70	(.30)
	Mean	3.80	4.90	(1.10)	3.80	4.45	(.65)
Pictures	Opposite Modality	6.20	6.10	(-0.10)	6.50	6.60	(0.10)
	Subordinate	6.00	6.00	(0.00)	5.30	6.70	(1.40)
	Geometric Figure	3.90	4.50	(0.60)	4.40	4.10	(-0.30)
	Verbal Representation	4.00	3.40	(-0.60)	3.00	3.70	(0.70)
	Mean	5.03	5.00	(- .03)	4.80	5.27	(.47)

Note.- The reader should keep in mind that a low automatizer does the task required on the automatization tests in a shorter period of time than does the high automatizer.

($\bar{X}$ = 4.15). In addition, this analysis yielded F (1,72) = 6.19, p < .05 for the effect associated with Sex of Subject; F (3,216) = 141.14, p < .01 for the effect due to Kind of Stimuli on the transfer task; and F (3,216) = 3.29, p < .05 for the interaction between Kind of Stimuli on the OL task and those on the transfer task. The major significance of the analyses described in this paragraph is to indicate only that where interactions of individual differences with treatments do occur, they are found with individual differences based on measures that appear to require some form of imaging. They do not occur with measures of cognitive processes or of personality (i.e., anxiety) behavioral tendencies.

It is interesting to note that the only personality or general intellectual ability measure to correlate with the transfer scores for female Ss (N = 104) was automatization. The correlations ranged from -.20 to -.25 (p < .025) for the four modes of presentation on the transfer task. However, the correlations between individual difference measures and transfer task measures for male subjects (N = 80) yielded ranges of .28 to .39 (p < ,01) for the vocabulary score; .14 to .33 (p < .05) for the Gottschaldt Figures Test; and .23 to .35 (p < .05) for the SAT scores (both verbal and math). Thus, automatization was the only influential correlate with performance for women Ss while only imagery and verbal factors were influential correlates with performance for male Ss. These differences suggest a possible reason for differences between the results of the present study and those of Stewart (1965).

Discussion

Based on the results from the total group of Ss, it is clear that learning pairs of items with pictures as stimuli is easier, that is, it takes fewer trials, than learning with words as stimuli. This finding provides a direct replication of a parallel finding from Stewart's (1965) study.

However, Stewart also found strong aptitude (Imagers) by treatments interactions. Thus, high imagers were found to learn pairs with pictures as stimulus elements more rapidly than with words as stimulus elements. Low imagers were found not to differ when the two kinds of stimuli were used. In addition, no significant differences were found between the two groups in transferring from pictures to words or vice versa, although there was a tendency for the high imagers to be hindered to a greater extent than low imagers by going from words to pictures.

In this regard, the present study suggests that automatization may be the influential behavioral tendency in distinguishing Ss performance when they are required to transfer from words to pictures compared to transferring from pictures to words. While differences between high and low imagers did not interact with treatments in the present study it was found that high-automatizers (those who take longer to perform the task) were particularly handicapped in transferring from words to pictures. Note that this process is clearly correlated with the process involved in performance on the automatization tests. Thus, automatization is measured by facility in labeling pictures without interference from other contextual stimuli; that is, they must go from pictures to words quickly and accurately. For example, in the Stroop color-name test, the S is required to read the word printed on the card rapidly. He is able to do so to the extent that he does not experience interference from the color of the print which is always different from the color-word. Similarly, on the automatization task, the S must name, as rapidly and as accurately as possible, the three objects pictured a total of 100 times on the card. He is able to do so in this task to the extent that he does not experience interference from the memory traces of the pictures already read. Thus, the significant differences in the performance of high and low-automatizers on the transfer task appears to reflect the cognitive operations that distinguish the two groups. In this respect, the present study replicates the parallel finding from Stewart's study.

The findings regarding the interaction between the original learning and transfer tasks indicate that transferring from words to pictures is more difficult than to transfer from pictures to words. Again this finding replicates one obtained by Stewart (1965). Developmental variables are undoubtedly implicated in explaining this result. In this culture, the S's typical experience is to label an object or picture; the opposite requirement is rare indeed. Moreover, these results imply that Ss form a "percept" before providing a label for the learning experience. This strategy is assumed to be a more dominant one for high-automatizers than it is for low-automatizers. The debilitating effects on the performance of high automatizers were especially noted when the Ss transferred from words in the original learning task to geometric representations and labels for geometric representations in the transfer task. The Ss did have somewhat more difficulty in transferring from pictures to verbal or pictorial representations than they did to subordinate categories or to the opposite modalities. However, they made nearly twice as many correct responses on the transfer task with the representation stimuli when the picture to word sequence was employed than when the word to picture sequence was employed.

The present study suggests clear replications of treatment effects related to the use of picture versus words in presenting stimuli and of treatment effects related to the strategies employed by Ss in processing learning materials. The implication of the differences in the processing strategies of high and low automatizers is an intriguing one and appears worthy of further investigation. The results suggest the need for careful analysis of processes involved in measures of individual differences and even, perhaps, the employment of task-specific measures if aptitude by treatment interactions are to be found. Our experience with the tasks involved in this experiment suggested that ceiling effects were approached if not reached in its conduct. For example, the four lists in the transfer task, comprised of a total of 32 items, could probably have been answered correctly on the second trial. For this reason only one trial had been used. A more sensitive design, perhaps coupled with more sensitive measures such as latency, is clearly indicated.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential aptitude tests. (Grades 8-13 and adults.) New York: Psychological Corporation, 1963.

Broverman, D. M., Klaiber, E. L., Kobayashi, Y., & Vogel, W. Roles of activation and inhibition in sex differences in cognitive abilities. Psychological Review, 1968, 75, 23-50.

College Entrance Examination Board Scholastic Aptitude Test. Princeton, N. J.: Educational Testing Service, 1962-1963.

Kuhlman, C. K. Visual imagery in children. Radcliffe College, Unpublished doctoral thesis, 1960.

Lockhart, R. S. Retrieval asymmetry in the recall of adjectives and nouns. Journal of Experimental Psychology, 1969, 79, 12-17.

Paivio, A. Learning of adjective-noun paired-associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370-379.

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, The University of Toronto, 1965.

Thurstone, L. L. A factorial study of perception. Chicago: University of Chicago Press, 1944.

Thurstone, L. L., & Jeffrey, T. G. Space thinking (Flags). Chicago: Education-Industry Service, 1959.

Yuille, J. C., Paivio, A., & Lambert, W. E. Noun and adjective imagery and order in paired-associate learning by French and English subjects. Canadian Journal of Psychology, 1969, 23, 259-466.

The Effects of Rated Vividness and Imagery of Learning Materials On Learning and Recall of High- and Low-Imagers

Francis J. Di Vesta, G. Susan Gray, Gary M. Ingersoll,
Steven Ross and Phyllis Sunshine

The learning strategy that differentiates high and low imagers was examined, albeit indirectly, in this study. Previous studies permit the conclusion that recall of words is affected by the qualities of vividness (Tulving, McNulty and Ozier, 1965) or imagery (Paivio, 1970). Furthermore, the degree of subjective organization has been found to vary according to the vividness of the list. For example, Tulving, McNulty and Ozier (1965) comment, "But the fact that the recall of more vivid words was ... organized by subjects to a greater extent thatn that of less vivid words is compatible with the hypothesis that vividness or picturability is an important component of meaning of words that affects the ease with which words can be grouped into higher-order memory units" (p. 250). However, these authors caution that such explanations are often inadequate since experiments in which only levels of vividness are manipulated deal only with correlations between stimulus characteristics and learning. Thus, for example, in the Tulving, *et al*. study, the effect might have been due to concreteness (or abstractness) or to associative relationships among the words in a given list rather than to "picturability" or other imaginal properties of the stimuli.

PRECEDING PAGE BLANK

In order to examine whether some organizational process related to imagery was employed by learners in such tasks as those used by Tulving, et al. (1965), Stewart (1965) examined the interaction between ability to use imagery as measured by the Flags (Thurstone & Jeffrey, 1959) and Spatial Relations (Bennett, Seashore, & Wesman, 1963) tests and the characteristics of the material learned. Exactly the same lists described by Tulving, et al. (1965) were used. As in the earlier study, the concrete (vivid) words were recalled more readily than abstract words. More interesting was the finding by Stewart (1965) that with these same lists the material was increasingly organized by high imagers as the vividness ratings of the words in the lists were increased while there were no differences in the degree of organization among lists by the low-imagers.

Stewart's results provide one basis for making inferences about the kind of processes that are employed by learners when approaching different learning tasks. They suggest too, that some treatments are more effective for learners who are imagers than for learners who are non-imagers. Since vividness (or picturability) implies an imaginal characteristic of symbolic stimuli, it would appear that imaginal transformations of stimuli (that is, words into "pictures") are as relevant as other cognitive processes in organizing materials for storage in memory.

Because of its implications for understanding strategies for learning and recall as well as for investigations of aptitude by treatment interactions, the present study was conducted to determine the replicability of Stewart's study. Furthermore, after a careful review of the literature on aptitude by treatment interactions, Cronbach and

Snow (1969) conclude that investigators have had difficulty in obtaining replications of such interactions, thus providing an important justification for attempting a replication of Stewart's study. The present study extends the earlier one by incorporating not only the original lists of words but also a set of lists based on the norms published by Paivio, Yuille, and Madigan (1968). As in the earlier study, the purpose of the present one was to investigate the hypothesis that the learning and recall of high-imagers is significantly greater when learning concrete words than when learning abstract words but learning of low-imagers is not differentially related to the concreteness or abstractness of the stimuli.

Method

Design

The overall procedure consisted of Ss learning lists of words that were high, medium, and low in concreteness (i.e., either rated vividness or rated imagery). The lists were presented in counterbalanced order. Each S was presented four trials of each list via the study-free recall procedure. The sequence of trials was different for each S learning a given list. The important features of the design implied a 2 x 3 mixed analysis of variance in which there were two levels of the between-subjects factor (vividness or imagery lists) and three levels of the within-subjects factor (high, medium, or low concreteness). Some of the analyses of variance included two levels of individual differences (either high and low imagers or male and female Ss) and order of presentation of the lists as between subjects factors. Correlations were computed for relationships between imagery, verbal ability and clustering scores.

Subjects

A total of 219 students enrolled in an introductory educational psychology course served as Ss in this study. Of these 104 were males and 115 were females. All Ss participated voluntarily but received credit toward their final grade for such participation.

Materials

Two sets of three word lists were employed: one set was based on rated vividness (V) and the other on rated imagery (I). The three lists (high, medium, and low) of 16 words within each set differed in the degree of rated-vividness or rated-imagery of their constituent words but were very nearly the same in terms of meaningfulness and Thorndike-Lorge (1944) frequency.

The V set of lists was identical to that described by Tulving, et al. (1965). In obtaining their values for vividness, they defined vividness as "the ease with which something could be pictured in the mind." Ratings were obtained on a 7-point scale with 1 corresponding to no image and 7 to extreme vividness. Meaningfulness of a word was obtained by ratings on a scale of 1 (corresponding to meaningless) to 7 (corresponding to extreme meaningfulness). The three lists in the V set are presented in Table 1.

The I set of lists was prepared from the Paivio, Yuille, and Madigan (1968) list of 1000 words rated for imagery and meaningfulness. The three lists in the I set were constructed in essentially the same manner as described for the V set. In the Paivio, et al. (1968) norms imagery was defined as the ease or difficulty with which a mental image

Table 1

The Experimental Lists of Words Based on Vividness Ratings*

	High Vivid (Concrete)			Medium Vivid			Low Vivid (Abstract)	
Word	V	M**	Word	V	M**	Word	V	M**
Apron	5.17	4.00	Abode	4.13	4.06	Buyer	2.94	3.93
Balloon	5.63	4.03	Bucket	4.50	3.88	Crisis	2.10	5.23
Bunny	5.23	3.93	Builder	4.00	4.28	Entry	2.08	3.71
Butler	5.79	3.79	Cargo	4.57	4.03	Founder	1.85	3.92
Cabbage	5.32	4.06	Fiber	3.96	3.50	Output	1.50	3.77
Camel	5.79	3.88	Hamlet	4.83	4.03	Patron	2.70	3.50
Chorus	5.63	4.43	Handful	4.36	3.43	Renown	1.33	3.94
Cigar	6.53	4.41	Madame	4.61	4.27	Routine	1.28	4.64
Circus	6.08	4.54	Pebble	4.76	4.15	Rover	2.11	3.30
Comet	5.81	3.96	Porter	4.64	3.79	Rumour	2.20	4.57
Granny	5.00	4.17	Pudding	4.61	4.05	Session	2.86	3.83
Jungle	6.00	4.36	Summit	4.77	4.46	Surplus	1.83	4.43
Lantern	5.43	4.32	Thicket	4.54	3.75	Tariff	1.50	3.61
Rainbow	6.13	4.94	Trainer	4.25	4.19	Topic	1.50	4.03
Runner	5.64	4.61	Veteran	4.07	4.40	Treason	1.50	4.70
Satin	5.42	3.62	Voter	4.32	5.15	Vigour	2.70	4.44
Mean	5.66	4.19	Mean	4.43	4.09	Mean	2.00	4.10

* From Table II, page 245, in Tulving, McNulty and Ozier (1965).

** Meaningfulness based on ratings.

was aroused by the word. Meaning was measured by Noble's (1952) production method. The lists of words in the I set with their associated imagery and meaningfulness values are presented in Table 2.

There were prepared 16 randomized trials of the words comprising each of the three lists within both the V and I sets, following the procedure described by Stewart (1965); that is, a given word did not occur in the same serial position, was not preceded by the same word, and was not followed by the same word more than once in all trials, eight of which were mirror images of the others. Each set of the 16 trials for a given list was placed on a single memory drum tape.

Procedure

The stimuli were presented to the Ss on a Stowe Memory Drum at a rate of one word per second. A given S was presented each of the three lists within a set: that is, he was presented all three versions, high (H), medium (M), and low (L), of either the I set or of the V set. The order of presentation of the lists within a set was randomized from one S to the next with the restriction that each of the possible combinations of list orders (LMH, LHM, MHL, MLH, HLM and HML) was equally represented at the completion of the study.

The Ss were administered four trials of a given list, the first of which was randomly selected from the 16 on the memory drum tape. The study-free-recall procedure, with standard instructions, was used. During the study period the S read aloud each of the 16 words in the list as they were presented. The end of the list was signalled by a dotted line after which the S was to write down all the words he could remember and in any order he chose. The free recall period was 90-secs.

Table 2

The Experimental Lists of Words Based on Imagery Ratings

	High Imagery (Concrete)			Medium Imagery			Low Imagery (Abstract)	
Word	I	M*	Word	I	M	Word	I	M
Acrobat	6.53	5.67	Anger	4.87	5.83	Belief	2.73	5.24
Alcohol	6.47	6.08	Assault	4.80	5.56	Crisis	3.43	5.44
Barrel	6.57	6.16	Barnacle	4.50	5.69	Edition	3.40	5.88
Bouquet	6.77	5.76	Belfry	4.43	6.25	Evidence	3.23	6.20
Candy	6.63	6.39	Blessing	4.43	6.19	Gender	2.90	5.41
Cigar	6.80	6.22	Chaos	4.57	5.88	Intellect	2.93	5.56
Factory	6.43	6.00	Clearness	4.77	5.64	Irony	2.83	5.24
Headlight	6.43	6.32	Comedy	4.90	6.08	Magnitude	2.50	5.68
Hotel	6.40	5.96	Humor	4.57	5.72	Malady	3.37	6.00
Hurdle	6.33	5.92	Mirage	4.97	5.63	Mastery	2.77	5.46
Jelly	6.40	6.00	Portal	5.10	5.63	Miracle	3.33	5.60
Piano	6.70	6.48	Reflex	4.73	5.88	Moral	3.17	6.44
Scarlet	6.37	5.80	Revolt	5.07	5.60	Origin	2.30	5.32
Slipper	6.47	6.04	Vacuum	4.77	5.94	Perjury	3.37	5.92
Steamer	6.53	6.32	Vapour	4.80	5.76	Satire	3.37	5.64
Tweezers	6.57	5.80	Victory	4.93	6.12	Welfare	3.17	6.16
Mean	6.52	6.05	Mean	4.76	5.84	Mean	3.05	5.70

* Meaningfulness based on Noble's (1952) production method.

long. This procedure was followed for all four presentations of a list following which there was a 2-min. rest period. The S was then administered the next series of items, either H, M, or L, depending on the condition to which he had been randomly assigned. The procedure was repeated until all three lists within a set had been presented.

Tests*

The following tests, fully described in another report (Di Vesta, Ingersoll, & Sunshine, 1971, in press) were administered to all Ss: the Space Thinking (Flags) test, the Spatial Relations test from the Differential Aptitude Test Battery, the Gottschaldt Figures test, the Stroop Color-Name test, the Automatization test, the Scholastic Aptitude (Math and Verbal) test, a vocabulary test, the Remote Associates test, a Reading Comprehension test, the Achievement Anxiety Scale and the Dogmatism Scale. The data were factor analyzed by the principal components method for the initial factorization. When the factors were rotated via the Varimax routine, three factors of relevance to this study were extracted: Verbal (as represented by the Scholastic Aptitude Test Verbal score), Imagery (as represented by the Spatial Relations test), and Automatization (as represented by the Stroop Color-Name). Raw scores for each test were standardized. The standardized scores for each test saturated on a given factor were summed to obtain a factor score. The Imagery factor score was comprised of the Flags, Spatial Relations, and Gottschaldt Figures Tests. The Ss who attained the highest 25 scores were classified as low Imagers. The distributions for men and women were considered separately. The same procedure was employed for high and low Automatizers based on the Stroop Color-Name

*The tests are described fully in the report entitled "A Factor Analysis of Imagery Tests" by Di Vesta, Ingersoll, & Sunshine, presented elsewhere in this report.

test and Broverman's Automatization Test; and for high and low Verbal Ability based on the Scholastic Aptitude Test: Verbal Score, a Vocabulary test, a Reading Comprehension test, and the Remote Associates Test.

Results

Overall Analyses of Number of Correct Responses

The number of correct responses on the recall trials were analyzed initially by an overall mixed analysis of variance in which measures of dispositional traits were ignored. This analysis was conducted to examine and determine the gross effects of manipulated variables and to aid in decisions about pooling of data for subsequent analyses. There were three between subjects variables: Sex of Subject, Kind of Lists (Vividness or Imagery), and six Orders of Presentation of lists. The within subject variables were: Levels of Concreteness (high, medium and low) and four Trials.

This analysis yielded F (1,168) = 46.89, $p < .001$ for the effect due to Sex of Subject; F (2,336) = 64.74, $p < .001$ for the effect due to Levels of Concreteness; F (3,504) = 726.64, $p < .001$ for the effect due to Trials; and F (3,504) = 18.44, $p < .001$ for the effect due to the interaction of Sex of Subject and Trials. None of the other interactions were significant ($p > .05$). The complete summary table for this analysis is displayed in Table 3.

These results indicated that women Ss averaged more correct responses ($\bar{X}$ = 10.68) per trial than did the men ($\bar{X}$ = 7.30). The Ss averaged more correct responses per trial on the concrete list ($\bar{X}$ = 9.47), than on the medium concreteness ($\bar{X}$ = 9.15) or the abstract ($\bar{X}$ = 8.35) lists. Average numbers of correct responses were 6.05 for the first,

Table 3

Summary of Overall Analysis of Variance of Correct Responses

Between Subjects	MS	df	F	p
Sex (A)	6444.14	1	46.89	<.001
Lists (B)	144.50	1	1.04	
Orders (C)	8.94	5	0.06	
A x B	6.15	1	0.04	
A x C	12.99	5	0.09	
B x C	21.91	5	0.16	
A x B x C	8.67	5	0.06	
Error (b)	139.55	168		
Within Subjects				
Levels (D)	257.05	2	64.74	<.001
A x D	2.09	2	0.53	
B x D	8.54	2	2.15	
C x D	6.93	10	1.75	<.10
A x B x D	6.93	2	1.75	
A x C x D	2.28	10	0.56	
B x C x D	3.20	10	0.81	
A x B x C x D	4.19	10	1.06	
Error (w)	3.97	336		
Trials (E)	2666.63	3	726.64	<.001
A x E	67.66	3	18.44	<.001
B x E	1.99	3	0.54	
C x E	1.85	15	0.50	
A x B x E	.74	3	0.20	
A x C x E	1.26	15	0.34	
B x C x E	2.23	15	0.61	
A x B x C x E	2.60	15	0.71	
Error (w)	3.67	504		
D x E	1.18	6	0.71	
A x D x E	.96	6	0.58	
B x D x E	3.15	6	1.89	<.10
C x D x E	2.17	30	1.30	
A x B x D x E	1.01	6	0.60	
A x C x D x E	1.34	30	0.81	
B x C x D x E	1.74	30	1.05	
A x B x C x D x E	1.13	30	0.68	
Error (w)	1.67	1008		

8.79 for the second, 10.16 for the third, and 10.95 for the fourth trials. The Sex by Trials interaction indicated that the women Ss learned the list more rapidly than men. Additional descriptions of the sex differences in learning these tasks will be provided in the analyses of individual differences.

The findings from this analysis are in substantial agreement with those obtained by Stewart (1965) and Tulving, et al. (1965). However, procedural differences between the ones used in those investigations and the present study should be noted as follows: In the present study four trials were used instead of eight, both men and women were used as Ss rather than only women, and lists varied in terms of imagery as well as vividness were employed. Despite these differences the finding that concrete lists are learned more readily than abstract lists was clearly replicated and performance over trials closely approximated the performance of Tulving, et al.'s (1965) Ss at the end of four trials. In addition, it was found that women learn the task more readily than men and that the Tulving, et al. (1965) lists produce essentially the same results as the lists based on the norms published by Paivio, et al. (1968). Unlike the earlier studies, practice effects or learning-to-learn were not observed in the present study. The reason for this difference may be due partly to the confounding of Lists and Orders and partly to the use of fewer trials than in the earlier studies. However, it should be noted that Stewart, who also confounded orders and lists found only a very small effect due to practice; her Ss averaged, over all trials: $\bar{X}$ = 13.02 for the first list learned, $\bar{X}$ = 13.44 words for the second list; and $\bar{X}$ = 13.56 for the third list.

Individual Differences in Imagery Related to Performance

The subsequent analyses of the number of correct responses were made by mixed analyses of variance with individual differences in Imagery (high and low) as the between variable and Trials and Levels of Concreteness (H, M, and L) as the within variables. Since the main effect of Sex as a factor and the interaction of Sex by Trials was significant in the initial analyses, separate analyses were made for men and women. Conversely, because there were no significant differences due to Kind of Lists in the previous analyses, the data for Ss administered the V or I lists were pooled for the present analyses. (See Tables 4 and 5.)

The analysis of variance of data for the men yielded no significant differences ($p > .10$) for the main effect of Imagery; F (2,96) = 16.65, $p < .001$, for the effect due to Levels of Concreteness; and F (3,144) = 431.57, $p < .001$ for the effect related to Trials. The effect of primary concern in this study, however, is that due to the interaction between Imagers and Level of Concreteness which yielded F (2,96) = 2.47, $p < .10 > .05$. The data comprising this interaction are presented for men and women Ss, separately, in Figure 1, and for the pooled groups in Figure 2.

The analysis of data for women yielded F (1,48) = 3.11, $p < .10$ for the effect due to Imagery levels. The main effects due to Levels of Concreteness and to Trials were significant ($p < .001$) as in the previous analyses. The effect due to the interaction of Imagery by Trials yielded F (3,144) = 4.71, $p < .01$, the data for which are presented in Figure 3. This ordinal interaction is identical to the one obtained by Stewart (1965) in all essential respects. None of the other interactions were significant ($p > .10$) in this analysis.

Table 4

Summary Table for Analysis of Variance of Individual Differences: Female

		Individual Difference Variable							
		Imagery		Automatization		Verbal		Anxiety	
	df	MS	F	MS	F	MS	F	MS	F
Between Groups									
Individual Difference (A)	1	65.34	3.11[a]	56.43	2.41	5.80	.262	17.68	0.85
Error (b)	48	21.03		23.39		22.18		20.78	
Within Subjects									
Levels (B)	2	45.50	12.62[c]	51.13	12.50[c]	51.81	12.25[c]	25.90	7.00
A x B	2	4.34	1.13	12.05	2.93[a]	6.59	1.56	5.13	1.39
Error (w_1)	96	3.84		4.09		4.23		3.70	
Trials (C)	3	919.23	569.31	958.92	376.47[c]	1022.64	483.26[c]	060.04	459.30[c]
A x C	3	7.60	4.71[b]	1.32	0.52	3.09	1.46	.99	0.47
Error (w_2)	144	1.61		2.55		2.12		2.11	
B x C	6	1.45	0.83	1.40	0.77	1.94	1.05	2.55	1.46
A x B x C	6	3.01	1.78	1.35	0.74	1.83	1.00	1.05	0.60
Error (w_3)(	288	1.74		1.83		1.85		1.74	

[a] $p < .10 > .05$

[b] $p < .01$

[c] $p < .001$

Table 5

Summary Table for Analysis of Variance of Correct Responses: Male

		Individual Difference Variable							
		Imagery		Automatization		Verbal		Anxiety	
	df	MS	F	MS	F	MS	F	MS	F
Between Groups									
Individual Difference (A)	1	8.88	.445	135.37	6.95[a]	24.40	1.15	105.84	5.71[a]
Error (b)	48	19.94		19.47		21.25		18.52	
Within Subjects									
Levels (B)	2	73.75	16.85[c]	49.01	10.20[c]	113.29	26.12[c]	74.48	18.72[c]
A x B	2	10.80	2.47[a]	2.86	0.59	8.21	1.89	4.58	1.15
Error (w_1)	96	4.37		4.80		4.34		3.98	
Trials (C)	3	817.69	431.57[c]	846.92	495.58[c]	836.61	399.71[c]	848.32	429.43[c]
A x C	3	.006	0.03	3.19	1.87	1.40	.67	1.36	0.69
Error (w_2)	144	1.89		1.71		2.09		1.98	
B x C	6	2.18	1.17	1.21	.62	2.44	1.12	2.63	1.62
A x B x C	6	1.77	0.95	1.70	.87	1.40	.64	2.34	1.44
Error (w_3)	288	1.87		2.00		2.19		1.62	

[a] $p < .10 > .05$

[b] $p < .05$

[c] $p < .001$

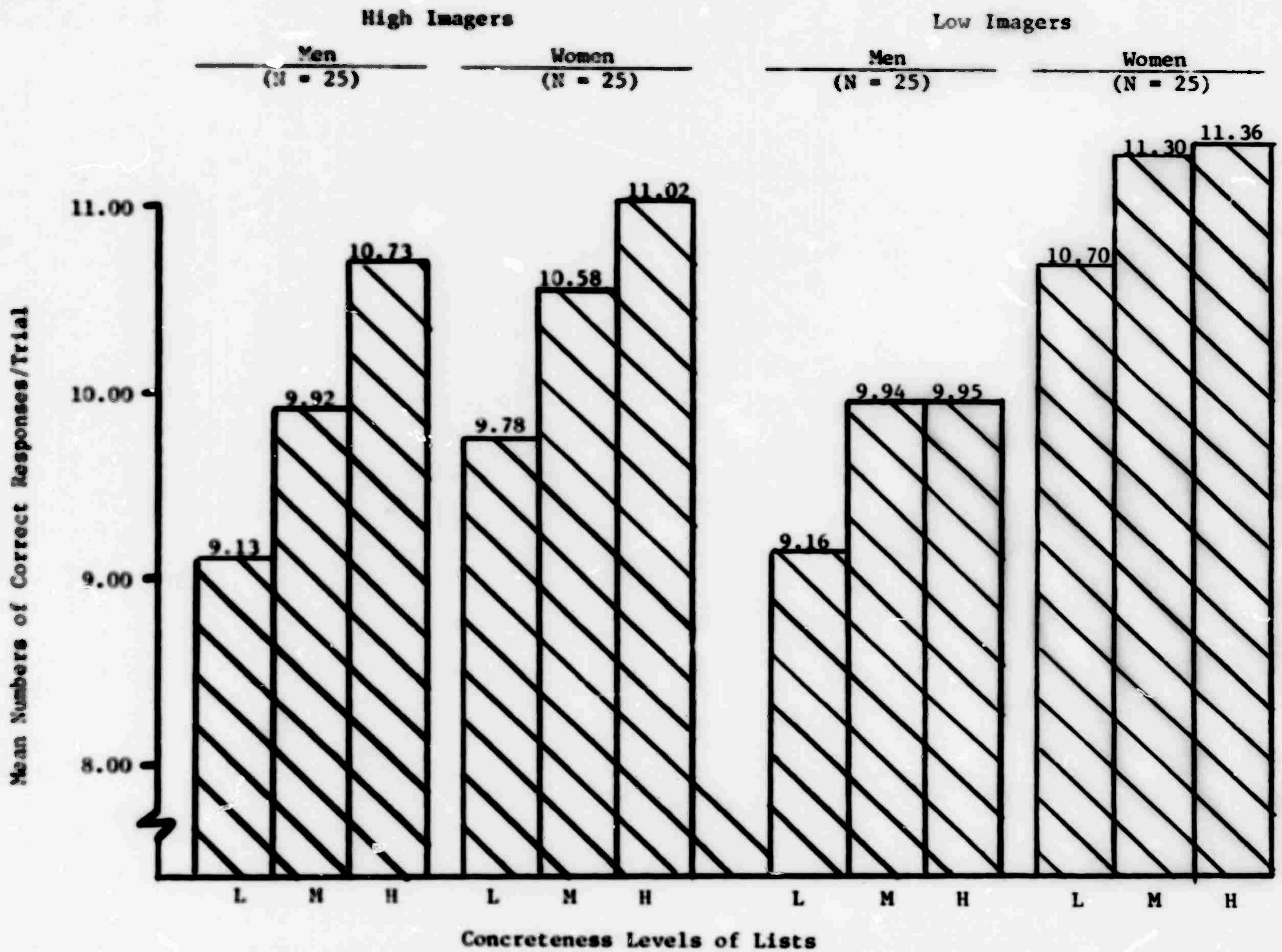

Figure 1. Mean numbers of correct responses made by men and women high and low imagers on lists comprised of low (L), medium (M), and high (H) concreteness words.

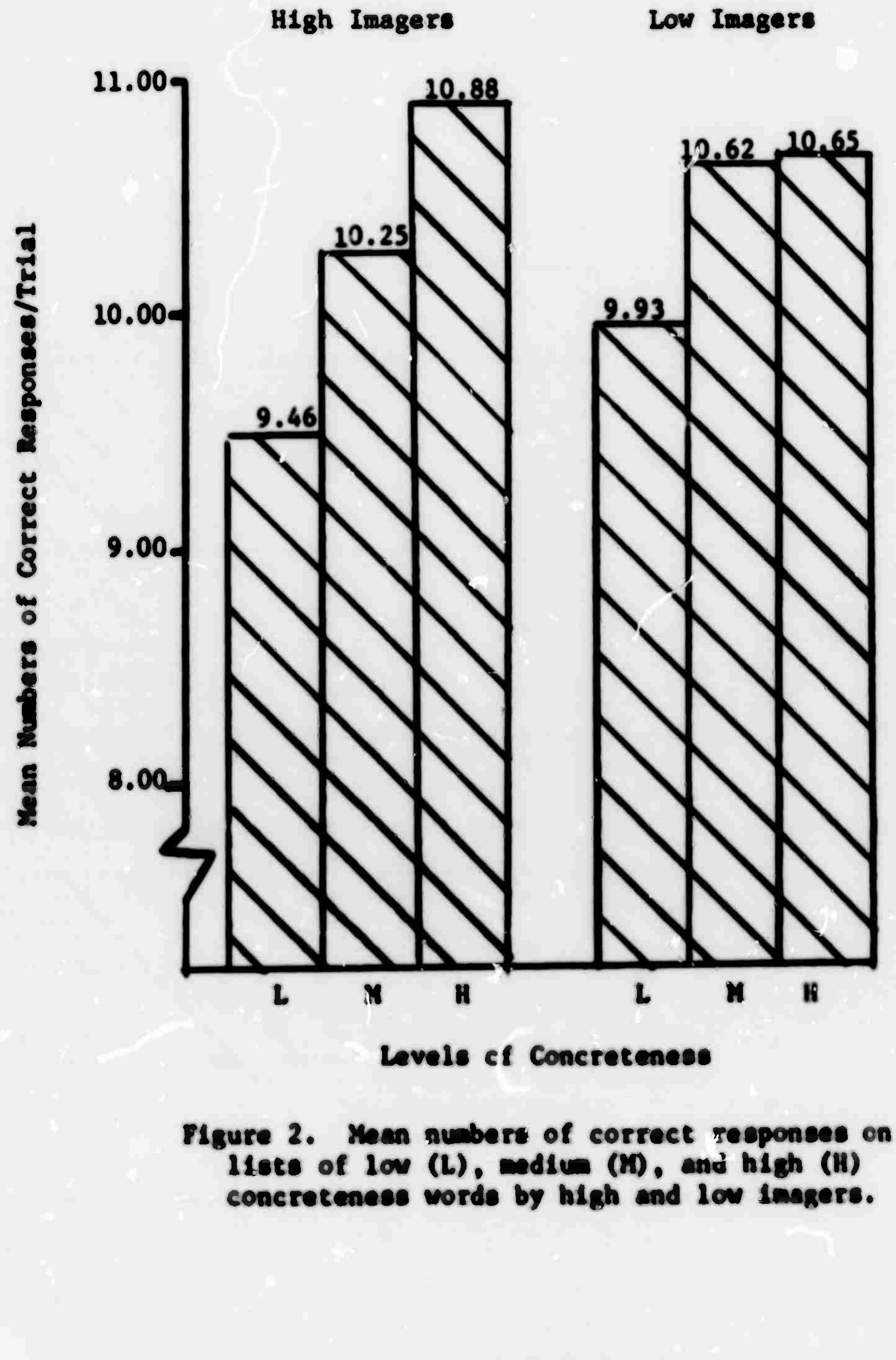

Figure 2. Mean numbers of correct responses on lists of low (L), medium (M), and high (H) concreteness words by high and low imagers.

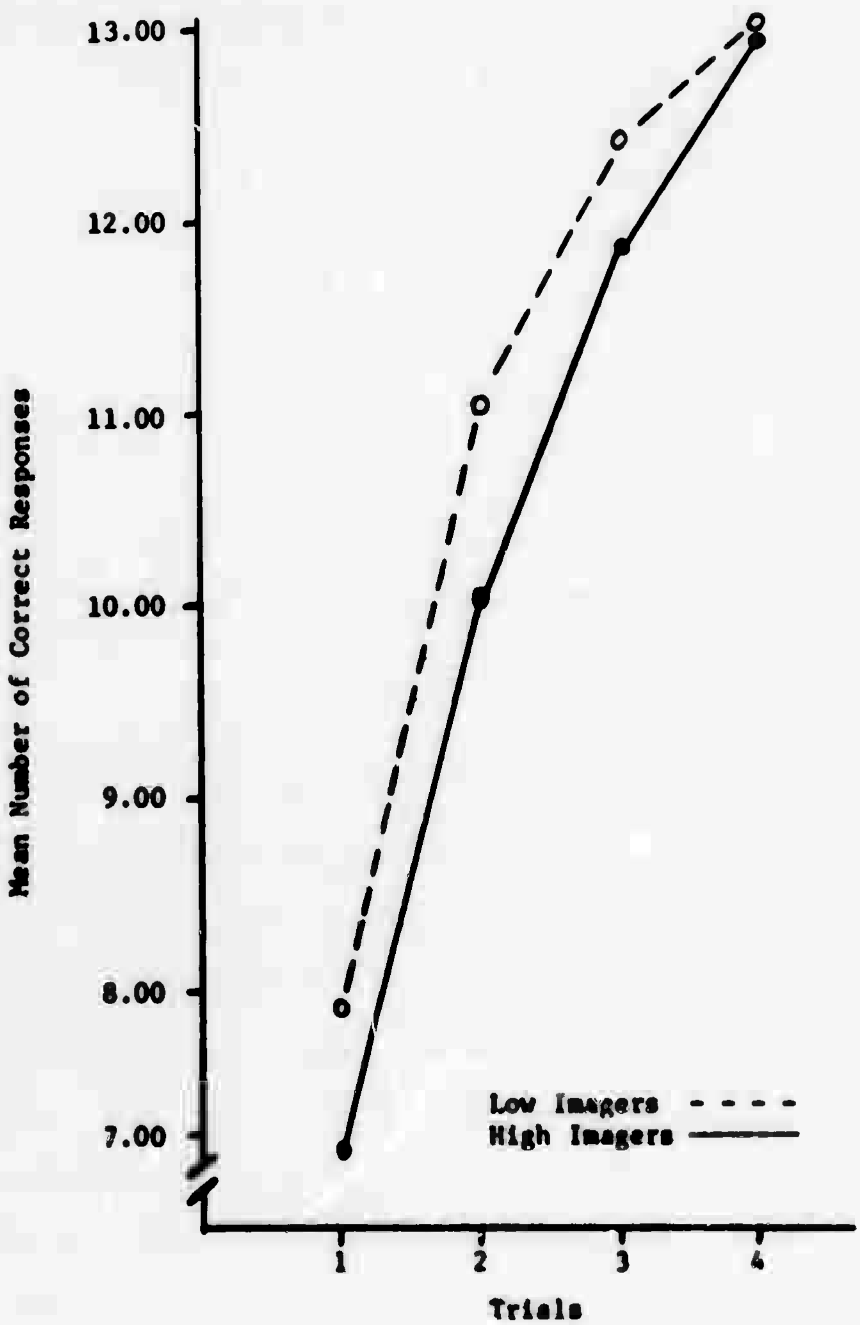

Figure 3. Mean number of correct responses for high- and low-imagers over trials for women subjects only.

The differences between Imagery levels, though not significant is interesting because the low imagers averaged more correct responses ($\bar{X} = 11.12$) than did the high imagers ($\bar{X} = 10.46$) over all trials. The differential performance of these two groups is more analytically reflected in the Imagery by Trials interaction. In this interaction ($\underline{p} < .01$) the high imagers perform at a much lower level ($\bar{X} = 6.85$) than the low imagers ($\bar{X} = 7.92$) on the first trial but they perform at about the same level ($\bar{X} = 12.94$) as the low imagers ($\bar{X} = 13.04$) on the fourth trial. Although the interaction between Imagers and Levels of Concreteness for the male <u>Ss</u> was not significant ($\underline{p} > .10$), their performance was very much like the women's performance and so have been presented in Figure 1 for purposes of comparison.

In order to provide a more direct comparison with the results from Stewart's (1965) study, another analysis was conducted identical to that described immediately above except that the high and low imagers were selected on the basis of only the Flags and Spatial Relations tests. The raw scores from each test were standardized and the two standardized scores for each <u>S</u> were then averaged. The women with the top 25 ranks on these scores comprised the group of high imagers and those with the bottom 25 ranks comprised the group of low imagers.

This analysis yielded significant differences ($\underline{p} < .001$) for the main effect due to Levels of Concreteness and to Trials. The effect due to the interaction between Imagery and Levels of Concreteness yielded $\underline{F}\ (2,96) = 2.41$, $\underline{p} < .10$ and that due to the interaction of Imagery and Trials yielded $\underline{F}\ (3,144) = 3.27$, $\underline{p} < .05$. These effects were essentially the same as displayed in Figure 1, 2, and 3.

The Imagery by Levels of Concreteness interaction, in this analysis where only the Flags and Spatial Relations tests were used to identify high and low imagers, indicated that women high Imagers obtained an average number of 9.85, 10.98 and 11.26 correct responses per trial for the low, medium, and high concreteness lists, respectively, while the women low imagers obtained an average of 10.40, 11.10 and 10.96 correct responses per trial for the low, medium, and high concreteness lists, respectively. As can be noted in Figure 1, the previous analyses in which factor scores were used yielded identical trends, though the interaction was not significant. As with data obtained from men Ss, these findings imply that high imagers benefit by increases in vividness or concreteness of the stimulus material where a strategy involving imaginal processes are employed. When concrete stimulus material is used, high imagers perform as adequately as the low imagers. The latter, on the other hand, perform with relatively equal efficiency on all tasks, though slightly less so on the tasks involving concrete materials; that is, the low imagers strategy for learning was relatively unaffected (unrelated) by the vividness of the task materials.

The performance differences between high and low imagers take on increased importance when compared with similar analyses based on other individual differences. Thus, identical analyses to those for imagery were made with the individual difference variables based on automatization, anxiety, and verbal scores. Significant differences ($p < .05$) were found for the main effects of automatization and anxiety based on the analysis of data for male Ss only. None of the other main effects (that is, the verbal factor for data based on the responses of the men, and automatization, anxiety, and verbal individual differences for data based on the responses of the women) were significant (p > .05). However, the main effects associated with Trials and Levels of Concreteness were, as might be expected, significant (p < .01) in every analysis. Among all possible interactions in these analyses only that between Automatization and Levels of Concreteness, based on the data for women Ss approached significance (p > .05 < .10). This interaction was very similar to that reported above for imagery differences and seems to be reasonable since the measures of Automatization were comprised of tasks involving visual materials.

Effect of Imagery-ability on Organization During Recall*

An analysis was made of the free-recall data on Trials 1 through 4 inclusive of the number of intertrial repetitions (ITR) and the number of correct responses common to trial N and trial N + 1 (ITC). The ITR's were computed following the procedure described by Bousfield and Bousfield (1966) and corrected by subtracting the expected ITR's from the observed ITR's. The ITC's were computed simply by counting the number of correct responses on trial N that were also on Trial N + 1. The number of new responses on trial N + 1 compared to trial N (ITN) were also computed by simple count. Finally, a sequential consistency (SC) score, which is an ITR score based on ratios, was computed according to a description by Fagan (1968).

These data were analyzed initially for possible differences due to list. Since there wre no significant list differences for either measure the lists were ignored in all subsequent analyses. Nevertheless, the two lists were equally represented in all conditions of the analyses described below. Although scores from all measures were analyzed, only

* The authors are indebted to Professor Susan Rosner at the University of Iowa, for the loan of her program for computing the several clustering scores indicated in the heading of Table 6 and Table 7.

the data for the ITR and ITC yielded trends of interest. Accordingly, these are the only analyses summarized here though the data for all scores are summarized in the accompanying tables of correlations to be presented below.

The scores derived from the two measures were analyzed in separate, mixed analyses of variance in which the between-subjects variables were Sex of subject and Imagery-ability. There were 25 Ss of each sex represented in the high imagery-ability and low imagery-ability groups, respectively. The within-subjects variables were Concreteness (rated imagery) of lists, and Trials compared (i.e., Trial 1 vs. Trial 2, Trial 2 vs. Trial 3, and Trial 3 vs. Trial 4).

The analysis of the ITR data yielded F (2,192) = 5.90, p = .003 for the effect due to Concreteness of list, and F (2,192) = 9.34, $p < .001$ for the effect due to Trials. The analysis of the ITC data yielded F (1,96) = 11.08, $p < .001$ for the effect related to Sex of subject; F (2,192) = 23.26, $p < .001$ for the effect due to Concreteness of List; F (2,192) = 767.31, $p < .01$ for the effect related to comparison among Trials; and F (2,192) = 4.67, $p < .01$ for the effect due to the interection of Trials by Sex of Subject. None of the other main effects of interactions were significant in either analyses. In general, these data indicate, especially for the ITC scores, that women Ss organized the words ($\bar{X}$ = 8.36) more than did men ($\bar{X}$ = 7.34); that there is more organization of lists with high concrete words ($\bar{X}$ = 8.45) than of lists with medium concreteness ($\bar{X}$ = 7.97), or low concreteness ($\bar{X}$ = 7.14); and there were fewer responses in common to Trials 1 and 2 ($\bar{X}$ = 5.65), than there were on Trials 2 and 3 ($\bar{X}$ = 8.20) or on Trials 3 and 4 ($\bar{X}$ = 9.70).

Concreteness clearly influences the ease with which a list is subjectively organized but the analyses failed to yield a significant interaction with imagery-ability except for a marginally significant ($p < .10$) interaction between imagery-ability and concreteness of lists when the ITC score was used as the dependent variable (see Figure 4). Accordingly, it was decided to compute correlations between imagery-ability scores and the several clustering measures separately for each of the lists representing different levels of concreteness. The entire sample of Ss ($N = 219$) was used in this analysis, the results of which are summarized in Table 6. The correlations computed separately for men and women Ss are also shown in that table.

As can be seen, there were significant correlations between imagery-ability and ITR scores for recall on the last two trials of the low-concreteness list ($r = -.14$, $p < .05$) and the first two trials of the high-concreteness list ($r = -.15$, $p < .05$). Although both are low correlations, they indicate that low-imagers achieve higher ITR's than do high imagers.

It is important, of course, to determine whether the relationships described above are confounded with some other ability. Unfortunately, it would be impossible to ferret out all such possibilities, but the most important one appeared to be the relationship between organization during recall and verbal (or general intelligence) ability. Accordingly, parallel correlations to those shown in Table 6 were computed by employing scores based on the tests comprising the verbal factor. These correlations are summarized in Table 7. None of these correlations was significant at the .05 level confidence, for data based on the total group the relationship between verbal ability and organization approached

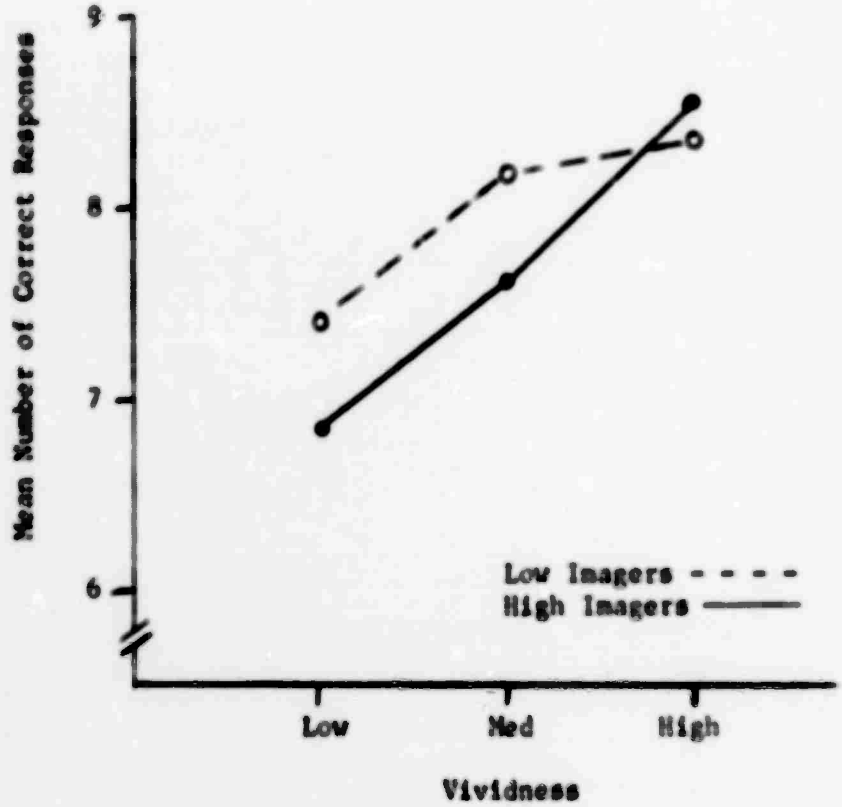

Figure 4. Mean number of correct words common to adjacent trials as a function of imagery-ability and list vividness.

Table 6

Correlations Between Measures of Organization in Recall and Imagery-Ability for Males and Females

List Imagery and Sex of Subject	Intertrial Repetitions (ITR)			Sequential Consistency			Intertrial Correct Responses (ITC)			Intertrial New Responses (ITN)		
	1-2	2-3	3-4	1-2	2-3	3-4	1-2	2-3	3-4	1-2	2-3	3-4
<u>High Imagery List</u>												
Males	-.05	-.06	-.01	-.10	-.06	-.04	-.04	-.17*	-.16	-.10	-.01	-.02
Females	-.24***	-.06	.05	-.19**	-.13	.02	-.18*	-.08	-.06	.04	12	.00
Total Group	-.14*	-.06	-.03	-.14**	-.09	-.01	-.08	-.02	.03	.05	.06	-.01
<u>Medium Imagery List</u>												
Males	-.04	.03	.08	.11	.01	.09	-.18*	-.09	-.03	.15	.01	.07
Females	-.09	-.08	-.03	-.06	-.04	-.02	-.28***	-.14	-.11	-.04	.06	.23**
Total Group	-.07	-.03	.01	.02	-.02	.03	-.23**	-.12*	-.08	-.08	.04	.16**
<u>Low Imagery List</u>												
Males	-.03	-.02	-.05	-.01	.03	-.06	-.03	-.07	-.00	-.01	-.04	-.03
Females	-.13	-.09	-.24***	-.09	-.10	-.21**	-.16*	-.25***	-.24***	-.08	.13	.30***
Total Group	-.09	-.05	-.15**	-.05	-.03	-.13**	-.11*	-.18**	-.14*	-.05	.05	.16**

* $p < .10$

** $p < .05$

*** $p < .01$

Table 7

Correlations Between Measures of Organization in Recall and Verbal-Ability for Males and Females

List Imagery and Sex of Subject	Intertrial Repetitions (ITR)			Sequential Consistency			Intertrial Correct Responses (ITC)			Intertrial New Responses (ITN)		
	1-2	2-3	3-4	1-2	2-3	3-4	1-2	2-3	3-4	1-2	2-3	3-4
<u>High Imagery List</u>												
Males	-.05	-.00	.11	.01	.02	.09	-.11	-.13	-.14	-.21	.18	.01
Females	-.03	-.01	-.11	.07	.00	-.09	.01	-.01	-.16*	.01	-.10	.08
Total Group	-.03	-.01	.02	.02	.00	.02	-.02	-.04	-.11*	-.10	.05	.02
<u>Medium Imagery List</u>												
Males	.05	-.00	-.08	-.00	.04	-.03	-.02	-.11	-.07	-.09	-.01	-.03
Females	-.11	-.08	-.21**	-.10	-.08	-.18**	.01	-.02	-.07	-.09	-.00	-.04
Total Group	.01	.01	-.04	.00	.07	.02	-.02	-.03	-.06	-.02	-.01	
<u>Low Imagery List</u>												
Males	.16	-.14	.08	.14	-.13	-.10	-.07	-.04	-.02	.11	.12	-.03
Females	-.04	.03	.03	-.05	.04	.05	-.24***	-.26***	-.29***	.01	.16	.28***
Total Group	.07	-.10	-.04	.06	-.08	-.01	-.10	-.08	-.07	.08	.14**	.09

* $p < .10$

** $p < .05$

*** $p < .01$

significance (p < .10) only on the comparison of ITC data based on the last two trials. Thus, there is at least partial assurrance that inferences concerning the use of imagery strategies are more accurate when discussed in the context of imagery ability than of verbal ability.

Similar analyses conducted separately for each sex indicate that the relationships described above are almost exclusively limited to female Ss. Furthermore, in many cases correlations of similar magnitude were found for relationships between verbal-ability and organization as for relationships between imagery-ability and organization although there were half again the latter comparisons as there were for the former. The correlation between verbal-ability and imagery ability for sub-groups was .16.

Discussion

It is clear that variations in the quality of concreteness attributable to words in a list are related directly to the facilitation of performance and recall in learning by the study-recall procedure. This finding is, of course, a replication of that obtained in earlier studies such as those by Stewart (1965), Tulving, et al. (1965), and Paivio (1965). It was also replicated within the present study where two sets of tasks were employed, each representing slightly different norms. Accordingly, although the stimulus materials were presented in a single modality, the higher the imagery-provoking value or picturability of the materials to be learned the easier they are to recall. This effect is analogous to the finding that pictures are recalled more easily than verbal materials when concreteness is held constant as found by Stewart (1965). Thus, beyond the mere replication of the relation between concreteness and performance, the importance of these data is that they imply a process by which Ss employ pictorial representations, similar to that suggested by the term "ikonic-imagery," as a strategy for encoding certain types of material.

The main effect attributable to sex differences is impressive only because it emerges so consistently in studies where tasks of the sort employed in the present one are used. The finding that the performance of women Ss was superior to that of men Ss is undoubtedly related to the effects of differential cultural experiences on the acquisition of knowledge and strategies for learning and recall by men and women. Such differences are reflected in differential performance on tests of general intellectual ability (Broverman, Klaiber, Kobayashi, & Vogel, 1968).

In this study there was no evidence for learning to learn. Previous evidence in similar investigations has been controversial. Both Stewart (1965) and Tulving, et al. (1965) obtained increments in performance over lists which they reasoned were due to learning sets. Dallett (1963), too, obtained the effect in his studies. Murdock (1960), on the other hand, concluded from his independent investigations, that neither learning to learn or warm-up effects were to be observed in multi-trial free recall learning tasks. In reviewing their results, Tulving, et al. (1965) suggest that, perhaps, learning to learn occurs only where uncued free-recall, rather than cued recall as in Murdock's experiment, was employed. However, this explanation cannot be applied to the present experiment since cues were not used in the recall phase. Learning to learn would appear to be a theoretically reasonable outcome of the present procedure. Nevertheless, the findings from the present study together with the fact that Stewart (1965) obtained only very small differences in her study, suggest that the effect must be considered a fragile one at best under the conditions of this experiment.

There were two interactions between Imagery and treatments that were of importance in this study. First, low imagers perform better over all trials than do high imagers. This observation replicates a comparable one observed by Stewart (1965). However, in the present study, there was an ordinal interaction between imagers and trials. Accordingly, while low imagers perform significantly better than high imagers on the first two trials their performances do not differ significantly on the last two trials. Apparently, high imagers employ the early stages of learning for identifying means by which their preferred strategy of imagery can be implemented in that task. A parallel finding by Paivio (1969; Ernest & Paivio, 1969; Ernest & Paivio, 1971) indicates that individual differences in imagery were predictive of incidental memory, indicating differences in learning strategies of high and low imagers. Thus, he found that high imagers recalled more incidental components of a compound stimulus or response item than did low imagers but the two groups did not differ in intentional learning. The evidence accumulating to date suggests that low imagers and high imagers employ quite different strategies in studying and recall with the consequent effect on rate of learning and on the amount and nature of what is recalled.

The second interaction of importance was that between Imagers and levels of concreteness represented by words in a given list. A series of t-tests among means indicated that high imagers had difficulty in learning a list of low concreteness (i.e., of abstract words). However, their performance increased significantly with increases in concreteness. Again, this finding replicates one obtained by Stewart (1965). Furthermore, Ernest & Paivio (1969) also reports similar sex differences, finding better incidental memory for high imagery females but not for males. As Paivio (1970) suggests, "It is too early to say what this might mean, but developmental factors certainly must be implicated."

Of interest in the study of aptitude by treatment interactions is the finding that organization is not manifest until the last trials on the more difficult (abstract) list and is manifested on the first trials of the easier (concrete) list. However, the implications of each differ: they suggest that there are more or less capricious attempts at organization on the first trials of the abstract list by both groups, while both groups (i.e., high and low imagers) organize the concrete list about equally after the first two trials. The correlations based on ITC data demonstrate differences in the strategies employed by the two groups somewhat more definitively. Thus, high imagers achieve significantly ($p < .10$) less organization than do low imagers on all comparisons of adjacent trials of the lists comprised of words with low-rated imagery, and on the comparison of the adjacent trials of the first three trials of the lists comprised of medium rated imagery. The degree of imagery-ability is unrelated to performance on any of the trial comparisons for the lists comprised of words rated high on imagery.

These data suggest that the high imagers may attempt to employ an inappropriate strategy for the abstract words thus hindering their performance on the low rated-imagery lists of words and to a lesser extent on the medium imagery words. On the other hand, the imagery strategy may be as appropriate for organizing the high imagery list as is any other strategy employed by Ss with low imagery ability. Or, perhaps, both groups employ the same strategy when learning lists of concrete words. In either case, the correlations are not significantly different from zero for the concrete list. In general, the correlation data lead to the conclusion that while imagery ability is not the exclusive factor involved in organization in recall, it is no less important than verbal ability.

These data reported here, together with those from other studies, further suggest that the preferred method of high imagers for encoding materials (that is, imaging) is inappropriate for low-concreteness words thus impairing their performance. However, the method is an efficient one when applied to materials that represent concrete referents or that can be imagined (pictured) easily. Accordingly, their (i.e., high imagers) performance improved in correspondence with increases in concreteness. The low imagers, who are hypothesized to employ other learning strategies, were relatively unaffected by changed in concreteness.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential aptitude tests. (Grades 8-13 and adults). New York: Psychological Corporation, 1963.

Bousfield, A. K., & Bousfield, W. A. Measurement of clustering and of sequential constancies in repeated free recall. Psychological Reports, 1966, 19, 935-942.

Broverman, D. M., Klaiber, E. L., Kobayashi, Y., & Vogel, W. Roles of activation and inhibition in sex differences in cognitive abilities. Psychological Review, 1968, 75, 23-50.

Cronbach, L. J., & Snow, R. E. Individual differences in learning ability as a function of instructional variables. Final Report, U. S. Office of Education, Contract No. OEC 4-6-061269-1217. Stanford, California: School of Education, Stanford University, 1969.

Dallett, K. M. Practice effects in free and ordered recall. Journal of Experimental Psychology, 1963, 66, 65-71.

Di Vesta, F. J., Ingersoll, G., & Sunshine, P. M. A factor analysis of imagery tests. Journal of Verbal Learning and Verbal Behavior, 1971, in press.

Ernest, C. H., & Paivio, A. Imagery and sex differences in incidental recall. (Research Bulletin No. 121) London, Ontario: Department of Psychology, University of Western Ontario, 1969.

Ernest, C. H., & Paivio, A. Imagery and verbal associative latencies as a function of imagery ability. Canadian Journal of Psychology, 1971, 25, 83-90.

Fagan, J. F. Measuring recall II: The ITR score expressed as a ratio. Psychonomic Science, 1968, 11, 205.

Mednick, S. A., & Mednick, M. T. Remote associates test. Boston: Houghton-Mifflin, 1967.

Murdock, B. B., Jr. The immediate retention of unrelated words. Journal of Experimental Psychology, 1960, 60, 222-234.

Noble, C. E. An analysis of meaning. Psychological Review. 1952, 59, 421-230.

Paivio, A. Abstractness, imagery, and meaningfulness in paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 32-38.

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, 76, No. 1, Part 2.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Thurstone, L. L., & Jeffrey, T. G. Space thinking (Flags). Chicago: Education-Industry Service, 1959.

Tulving, E., McNulty, J. A., & Ozier, M. Vividness of words and learning to learn in free recall learning. Canadian Journal of Psychology, 1965, 19, 242-252.

A Factor Analysis of Imagery Tests

Francis J. Di Vesta, Gary Ingersoll and Phyllis Sunshine

The mentalistic sounding construct "imagery" has gained renewed acceptance on several experimental fronts during the past decade and undeniably has led to a number of fruitful insights on learning processes. Despite its many definitions (Holt, 1964), recent investigators have found that to provide this concept with operational meaning, either by reference to the *S*'s behavior or by inference from objective descriptions of experimental conditions, was a relatively straightforward matter.

Several orientations in studies of imagery can be identified in the current literature. In one, represented by the work of Paivio (for example, see Paivio, 1970), the concern has been with the effects of stimulus characteristics, such as the perceived vividness or concreteness of events, usually language symbols, on the efficiency of learning processes. In another, the emphasis is on imagery as a *process* or *strategy* by which experiences become encoded or transformed for storage in, or later retrieval from long-term memory. Bower's (for example, see Bower, 1970) research is an illustration of this category of investigations. Within a third framework, imagery has been defined in terms of individual differences based on the subjects' reports of vividness of imaginal experiences (for example, Galton, 1880, 1883; Richardson, 1969) or on scores for objective tests (for examples, see Thurstone, 1944).

A potentially fruitful extension of the aforementioned orientations may exist within the aptitude by treatment interaction (ATI) framework proposed by Cronbach (1957) as a general methodology for blending the experimental and correlational methods. It was this orientation that provided the impetus for the present study. Examples of the application of ATI to the study of imagery are to be found in investigations described by Hollenberg (1970) and Stewart (1965). Only a few such studies have been conducted to the present. Nevertheless, it is becoming increasingly apparent (Paivio, 1970; Rohwer, 1970) that hypotheses related to the differential effects on performance of manipulated variables as they interact with differences in ability (or preference) to use ikonic imagery versus verbalization strategies in thinking are gaining attention.

An essential requirement for studies within the ATI orientation is a reliable and valid measure of imagery. Introspection was the basis for the earlier measures proposed by Galton and was retained in the more recent summary by Richardson (1969) who described the revised scales originally employed by Betts (1909) and Gordon (1949). A skeptical view of the reliability of self-reports of imagery and the consequent search for more objective measures prompted Hollenberg, in her investigations of visual imagery with children, and Stewart, in her studies with college students, to reject the self-report procedure and to employ spatial manipulations tests instead.

Both Hollenberg (1970) and Stewart (1965) made some further assumptions that influenced not only the choice of their tests but also the interpretation of the scores. They reasoned that individual differences in thinking by the use of images (visualizers), on the one hand, or by the use of language symbols (verbalizers), on the other, were the products of the individual's unique history of rewards and punishments for employing a given strategy. Permissiveness in child-training was believed to influence the continued use of imagery and of language habits related to imagery. Guidance in the use of symbols not directly based on perceptual similarities was reasoned to encourage the use of verbalizations in thinking. As the child begins to think through the use of symbols, ikonic imagery was assumed to fade; that is, verbalization is substituted for imagery. These assumptions led both investigators (Hollenberg, 1970; Stewart, 1965) to employ the spatial manipulations tests as though imagery was inversely related to verbal ability; that is, they assumed high scores represented imagers and low scores represented verbalizers. This inference was supported, in part, by the disordinal interactions found between treatments and aptitudes. Thus, high-imagery subjects (visualizers) tended to perform more effectively than low-imagery subjects on tasks hypothesized to favor ikonic mediation while low-imagery subjects (verbalizers) outperformed the high imagery subjects on tasks that favored verbal mediation.

This brief review suggests the hypotheses that (a) measures of imagery and verbal ability are independent (orthogonal), if not bi-polar, factors; and (b) measures of imagery based on introspection are different from and/or less reliable than measures based on objective tests. These hypotheses, of course, are directed toward an examination of the construct validity (Cronbach & Meehl, 1955) of imagery as an individual difference variable.

One means of testing these hypotheses is to determine the relationships between the three kinds (introspective reports of imagery, objectively defined tests of spatial manipulation, and tests of verbal ability) of tests to the performance of Ss on tasks assumed to be facilitated differentially by verbal and imaginal processes. This approach was the one employed by Hollenberg and Stewart. However, attempts to replicate Stewart's experiments in our laboratory were only partially successful. Accordingly, following Barrett (1953), the alternative procedure of testing these hypotheses via factor analytic procedures was used, and is the subject of this report. The underlying simple structure represented in several tests was examined in two separate studies. In the first study, the hypothesis related to the independence of verbal ability and imagery was examined; in the second study the relationship between introspective and objective tests of imagery was investigated.

Method

Subjects

There were 184 Ss in Study I and 232 Ss in Study II. All Ss were enrolled in an introductory course in educational psychology. Though participation was voluntary, Ss received credit toward their course grade for participating in the study. Most Ss had taken part in one or more experiments prior to enrollment in this course.

Materials

Study I. The battery of tests for Study I was purposely contrived to consist of at least two factors: One group of tests was hypothesized to depend primarily on perceptual skills or spatial manipulation. The second group consisted of tests related to verbal and general intellectual abilities. Another group, comprised of general personality variables, was included to prevent restrictions on the extraction of factors. The specific tests in this battery are described immediately below.

The Gottschaldt Figures Test as described by Thurstone (1944) consisted of 61 test items divided into five parts. There were 27 items in Part A, 7 items in Part B; 7 items in Part C; 10 items in Part D; and 10 items in Part E. Maximum time limits were two, one, three, four, and four minutes for each part, respectively. The score was the number of designs correctly traced within the time allowed.

The *Space Relations* test of the *Differential Aptitude Test Battery* (Bennett, Seashore, & Wesman, 1963) was employed as a measure of the ability to visualize the rotation of a picture or pattern in three dimensional space. The score for the total number of correct choices was obtained.

The *Space Thinking* (*Flags*) (Thurstone & Jeffrey, 1959) test was intended to measure an ability similar to that described for the *Space Relations* test. A time limit of 20 minutes was imposed. The score recorded was the number right minus the number wrong.

The *Stroop Color-Word Interference Test* (Stroop, 1935) was constructed according to standard procedures (for example, see Jensen and Rohwer, 1966; Thurstone, 1944). Three forms were administered to each *S*. In one the *S* read the names of colors printed in black. In a second, he named the colors of patches of color. The third version was the color-word interference task in which each word was printed in a color other than its color name. The total time for reading each version correctly was recorded.

The *Automatization Test* (Broverman, 1964; Broverman, Klaiber, Kobayashi, & Vogel, 1968) measured the *S*'s rate of naming three objects (tree, fly, and cup) repeated equally often on a card. There were depicted 110 representations of these objects on the card. The *S*'s score was the amount of time taken to name all objects correctly.

A *Vocabulary Test* was specifically devised for this study by modifying in multiple-choice form, several items selected from the *Henmon-Nelson Test of Mental Ability* (Lamke, Nelson, & Kelson, 1931-1960). This test consisted of nine verbal analogies, 17 vocabulary definitions or meanings, and four opposites.

The Scholastic Aptitude Test (SAT) of the College Entrance Examination Board (1962-1963) had been administered to all Ss prior to entrance to the university. Their verbal and mathematics scores on this test were obtained from their college records.

The Reading Comprehension test (Lindsay, Williams, & Peterman, 1969) was developed by the Student Affairs Research Office at the Pennsylvania State University. It consisted of 12 paragraphs, each of which was followed by two or more test items for each paragraph, for a total of 30 items. This test had been administered to the Ss during their Freshman year. Scores were obtained from their college records.

The Remote Associates Test (Mednick & Mednick, 1967) was administered as a test of ability to make mediating links in groups of words and, hence, was considered as a potential contributor to the verbal factor. It was administered with a 30-min. time limit. The score was the number of items answered correctly according to the key provided in the manual.

The personality measures consisted of the Achievement Anxiety Test (Alpert & Haber, 1960) from which debilitating anxiety and facilitating anxiety scores were obtained; the Tolerance for Ambiguity test devised by Budner (1963); and the Dogmatism Scale (Rokeach, 1960). These tests were devised, administered, and scored following descriptions provided in each of the references shown.

Study II. The test battery for Study II included the Space Thinking (Flags) Test; the Scholastic Aptitude Test (SAT); the Spatial Relations Test; and the Gottschaldt Hidden Figures Test, all of which have been described under Study I. In addition, the following were administered: the Digit Span test from the Wechsler Adults Intelligence Scale (Wechsler, 1955); the Tolerance for Ambiguity Scale (Budner, 1963); the Social Desirability Scale (Crowne, & Marlowe, 1964); the Memory-for-Designs Test (Graham & Kendall, 1960); the Betts Vividness of Imagery Scale (Betts, 1909); and the Gordon Test of Visual Imagery Control (Gordon, 1949). The latter two tests were employed as described in Richardson (1969). These tests, too, were devised, administered, and scored as described in the references cited.

Procedures

The tests for both studies were administered according to standardized instructions and procedures provided in manuals and references for the tests. There were, however, two exceptions: The items for the Digit Span Test were recorded and the items for the Memory-for-Designs tests were placed on 35-mm. slides, so that they could be presented to small groups rather than individually. The Stroop Color-Word Tests and the Automatization Test were administered individually to the S usually prior to or following participation in another experiment. All other tests were administered to Ss in small groups of 15-25 Ss which were monitored by two Es.

Results

In both Study I and Study II the basic data were the raw scores from the tests employed in each study. Pearson product-moment correlations among all scores within a study were calculated and then used in the principal components analysis for initial factorization. Six factors were extracted with eigenvalues greater than 1.00 in both studies. The factors extracted were rotated via the Varimax routine (Kaiser, 1958). Stability of the factor structure was achieved for the rotation of four factors in Study I and for three factors in Study II. The results of these analyses are described below.

Study 1. The first study was concerned with the question of whether imagery and verbal ability were separable individual difference characteristics or whether they were constituents of a more general intellectual ability. The means and standard deviations of scores from the 20 aptitude tests are summarized in Table 1 separately for males and females. The intercorrelations among all variables for all *Ss* (N = 184) are displayed in Table 2. Note that sex was included as a "score" ("one" was employed for females and "two" for males) in a manner similar to that employed in Thurstone's (1944) earlier studies of perception. The authors were aware of difficulties associated with this practice but in view of the similarity in results obtained via separate analyses for each sex, it was decided that the most parsimonius means of presenting descriptive data for the entire group was by the summary of the analysis based on all *Ss*.

The results of the analyses are shown in Table 3. The first factor is comprised of Reading Comprehension, the Scholastic Aptitude Test, the Vocabulary Test, and the Remote Associates Test. These measures appear to be clearly classifiable as *Verbal* or *Symbolic Imagery*.

The second factor is comprised of the Flags, Spatial Relations, Gottschaldt, and the SAT:Math tests scores. This factor, with the exception of the SAT:Math test (which has the lowest of the loadings) can be defined as *Ikonic Imagery*.

The third component extracted was named *Automatization*. It was comprised of the three forms of the Stroop Color-Word Test and the Automatization test. Although it had been assumed that these tests might have had large "imaginal" components it is clear that the factor is separately defined from Ikonic Imagery thereby supporting results

Table 1

Means and Standard Deviations for Females and Males

On Twenty Measures: Study I

Test	Females (N = 104)		Males (N = 80)	
	$\overline{X}$	SD	$\overline{X}$	SD
Flags	106.40	19.98	115.50	12.84
Spatial Relations	68.39	16.77	76.35	15.91
Gottschaldt Figures Test: I	16.59	7.43	18.41	7.82
Gottschaldt Figures Test: II	5.28	2.02	5.80	1.85
Gottschaldt Figures Test: III	4.62	1.98	5.19	1.77
Gottschaldt Figures Test: IV	4.59	2.64	5.85	2.94
Gottschaldt Figures Test: V	6.39	2.04	7.30	2.05
Stroop: Word Score (secs.)	38.02	4.94	40.41	4.75
Stroop: Color Score (secs.)	52.29	7.03	56.83	7.84
Stroop: Interference Score (secs.)	93.21	15.69	100.30	19.04
Automatization	53.06	7.81	54.40	6.70
SAT: Math	549.50	83.07	569.00	78.41
SAT: Verbal	527.30	91.24	500.80	78.81
Vocabulary	19.13	3.73	18.23	3.97
Remote Associates Test	13.40	5.06	13.59	4.51
Reading Comprehension	16.21	4.35	16.24	4.62
Debilitating Anxiety	27.59	6.48	25.29	4.92
Facilitating Anxiety	24.36	4.80	26.01	4.07
Tolerance of Ambiguity	41.35	8.83	42.48	8.76
Dogmatism	128.90	19.14	130.70	18.00

Table 2

Matrix of Intercorrelations Among Variables in Study I

Variables	SR	G:I	G:II	G:III	G:IV	G:V	S:W	S:C	S:CW	A	SATM	SATV	V	RAT	RC	DA	FA	TA	D	S
Flags	35	33	19	32	21	29	13	05	00	-01	24	-12	-10	06	10	-03	10	10	14	26
Spatial Relations		22	31	17	25	40	23	12	09	06	36	-04	04	08	02	02	-04	16	11	24
Gottschaldt: I			40	21	28	39	-06	04	-06	-13	24	11	05	-02	01	-11	11	10	08	12
Gottschaldt: II				41	38	47	02	00	-12	-10	38	22	06	12	-01	-07	07	07	00	13
Gottschaldt: III					55	44	09	-01	-08	-04	20	-01	-06	10	-05	-04	06	00	00	15
Gottschaldt: IV						59	04	01	-07	-03	29	10	09	15	01	-09	15	12	04	22
Gottschaldt: V							08	05	-04	00	37	13	07	14	08	-04	12	11	09	22
Stroop: Word								54	54	45	00	-29	-30	-05	-15	09	-11	04	11	24
Stroop: Color									71	60	05	-13	-16	-03	-11	04	-04	-01	11	29
Stroop:Color Word										59	-07	-21	-19	00	-13	04	00	-13	09	20
Automatization											-06	-12	-17	00	-15	10	-09	-12	05	09
SAT: Math												35	26	18	29	-14	17	03	01	12
SAT: Verbal													62	29	51	-24	20	-02	-16	-15
Vocabulary														33	48	-19	21	-11	-13	-12
RAT															11	-12	07	-03	-12	-02
Reading																-19	28	02	00	00
Deb. Anxiety																	-61	02	23	-19
Fac. Anxiety																		04	-14	18
Tol. of Ambiguity																			23	06
Dogmatism																				05
Sex																				

Table 3

Summary of Factor Loadings for Tests Associated with

Four Varimax Factors: Study I*

Test	I Verbal	II Imagery	III Automatization	IV Anxiety	h^2
			Factor		
Flags	-.12	.57	.08	.09	.42
Spatial Relations	.09	.58	.22	-.15	.46
Gottschaldt Figures Test: I	.03	.57	-.08	.08	.35
Gottschaldt Figures Test: II	.18	.67	-.06	-.04	.54
Gottschaldt Figures Test: III	-.06	.65	-.04	.03	.60
Gottschaldt Figures Test: IV	.10	.70	-.02	.07	.57
Gottschaldt Figures Test: V	.16	.77	.04	-.02	.63
Stroop Test: Word	-.21	.14	.73	-.08	.60
Stroop Test: Color	-.02	.04	.86	.00	.75
Stroop Test: Word/Color	-.07	-.11	.86	.03	.76
Automatization	.00	-.11	.78	-.10	.65
SAT: Math	.50	.49	.06	.04	.51
SAT: Verbal	.83	.02	-.17	.11	.73
Vocabulary	.81	-.05	-.17	.10	.70
Remote Associates Test	.50	.10	.06	.02	.32
Reading Comprehension	.63	-.00	-.10	.20	.61
Debilitating Anxiety	-.15	-.05	.04	-.85	.74
Facilitating Anxiety	.15	.11	-.03	.83	.74
Tolerance of Ambiguity	-.15	.26	-.12	-.06	.42
Dogmatism	-.15	.19	.10	-.35	.51
Sex**	-.19	.37	.34	.39	.46
Eigenvalues	2.12	3.78	3.35	1.49	

* N = 80 males and 104 females.

** 1 = Females; 2 = Males.

obtained by Broverman (1964). This factor refers to the <u>S</u>'s ability to perform simple repetitive tasks without being distracted by interfering influences, such as the general effect produced when the actual colors of the printing interfere with reading color-names.

The fourth factor obtained was <u>Anxiety</u>. As one might expect, this factor was comprised of the two scores, representing facilitating and debilitating anxiety, from the Achievement Anxiety Scale. This factor is interesting only because the two scores are differentially polarized on the factor, thereby providing a degree of validity for the constructs as hypothesized in the development of the scales (Alpert & Haber, 1960).

None of the loadings of the sex variable are high for any of the factors. Nevertheless, it is of interest to note the directions of the correlations: the Verbal factor was inversely correlated and the other factors were positively correlated with sex indicating that females tended to achieve higher scores on the Verbal factor than did males and males tended to achieve higher scores on the Automatization and Ikonic Imagery factors than did females. A similar tendency was noted by Thurstone (1944).

<u>Study II</u>. The analysis of data obtained for the second study was directed toward examining the validity of the notion that introspective reports and objective tests of imagery ability provide measures of the same individual difference characteristics. The means and standard deviations of scores for all tests in this study are summarized in Table 4.

These data are based on the entire pool of <u>Ss</u> for Study II. A comparison with Study I indicates the results from the two studies were within a standard deviation of one another for overlapping tests,

Table 4

Means and Standard Deviations for Scores

On Ten Tests: Study II*

Test	$\overline{X}$	SD
Social Desirability	14.65	5.06
Vividness of Imagery	196.94	29.68
Control of Imagery	29.86	6.79
Digit Span	9.26	2.05
Memory for Figures	16.47	3.71
SAT: Verbal	509.47	88.04
SAT: Math	552.13	86.86
Gottschaldt Figures Test	36.10	11.89
Flags	108.93	19.98
Spatial Relations	67.72	17.68

* N = 232

although in each case the mean scores for Ss in Study I were higher than those for Ss in Study II. Thus, for Studies I and II, respectively, the SAT:Verbal scores were 515.80 and 509.47, the SAT:Math scores were 558.00 and 552.13, the Flags Test scores were 110.40 and 108.93, the Gottschaldt Figures Test scores wre 39.65 and 36.10, and the Spatial Relations Test scores were 71.85 and 67.72.

The intercorrelations among all variables in Study II are presented in Table 5. The correlations between the same variables in Study I and Study II w[illegible] comparable except for the correlations between SAT:Verbal and the Spatial Relations test which were -.04 in Study I and .21 in Study II; and between SAT:Math and SAT:Verbal which were .35 and .54 in Study I and Study II, respectively.

The summary of the rotated factor matrix for Study II is shown in Table 6. The results presented there provide a clear reproduction of the Verbal and Imagery factors extracted in the factor analysis for Study I. As in Study I, women were found to be more facile in verbal than were the men and men were higher in imagery ability than were women. In addition, there was extracted a third factor described by the label Social Desirability. The constituents of this factor were the Control of Imagery Scale, the Vividness of Imagery Scale, and the Social Desirability Scale.

In summary, Study II provided a replication of the distinction between verbal and imagery abilities and, in addition, indicated that objective tests provide measures of abilities that may be quite different from those provided by introspective reports.

Table 5

Matrix of Intercorrelations Among Variables in Study II*

Variables	SD	VI	CI	DS	MFF	SATV	SATM	GFT	F	SR
Sex**	04	01	04	-11	00	12	-05	03	-22	-14
Social Desirability		29	20	-07	05	-12	-06	02	-03	-04
Vividness of Imagery			45	00	13	-06	00	06	-04	-04
Control of Imagery				03	07	09	14	09	-01	-08
Digit Span					02	22	18	02	05	07
Memory for Figures						15	28	26	24	38
SAT: Verbal							54	30	-06	21
SAT: Math								44	25	29
Gottschaldt Figures Test									28	34
Flags										37
Spatial Relations										

*Decimal points have been omitted in correlation coefficients.

**In this study men were assigned a score of one and women were assigned a score of two.

Table 6

Summary of Factor Loadings for Tests

Associated with Three Varimax Factors: Study II

Tests	Factor I Verbal	Factor II Imagery	Factor III Social Desirability	h^2
Flags	-.02	.78	.04	.61
Spatial Relations	.26	.70	.08	.56
Gottschaldt Figures Test	.51	.44	-.16	.48
Memory for Figures	.26	.53	-.21	.40
Control of Imagery	.17	-.08	-.75	.59
Vividness of Imagery	-.04	.01	-.81	.66
Social Desirability	-.20	.03	-.62	.43
SAT: Math	.75	.32	-.06	.67
SAT: Verbal	.86	-.07	.04	.76
Digit Span	.37	-.01	.09	.14
Sex*	.26	-.47	-.12	.31
Eigenvalue	1.40	2.52	1.70	

* In this study men were assigned a score of one and women were assigned a score of two.

Discussion

The results of this study clearly confirm the stability of imagery as an individual difference variable. In large part, the tests comprising this factor have no apparent dependence on obvious meaningful associations. An analysis of manipulations required for each set of tasks would imply that they could only be accomplished with minimal benefit from verbal associations, labeling, or conceptualizations. Furthermore, it appears that successful performance on these tasks required those mental abilities with imaginal rather than symbolic properties.

The processes, presumably indexed by the imagery factor can be inferred from an analysis of the kinds of skills required to perform the tests which saturate this factor in both studies. The ability to hold a percept in memory long enough to work with it was measured by the Memory-for-Designs test. The Flags and Spatial Relations tests require the S not only to hold the percept in memory but to rotate it or to unfold it in various ways, that is, to "view" it from many perspectives. However, these are relatively primitive abilities. The measure of the higher forms of imagery would require a task in which the figure would be changed or reorganized. To some extent this function was served by the Gottschaldt Figures Test. In these tasks, the critical figure to be identified was camouflaged by extraneous lines. In order to perceive the hidden, less obvious stimulus relationship the initial more obvious percepts had to be subdued. Thus, the underlying process in the Gottschaldt test appears to be one of not only holding the stimulus in memory but also of restructuring the percept.

The present factor analysis, of course, only implies that the Flags, Spatial Relations, Memory-for-Designs and Gottschaldt tests comprise a cluster of tasks that presumably require some common process. It does not, except by inference, indicate whether imaginal, visualization, perceptual, or some other process underlies performance of the tasks involved. An integration of a finding from Barratt's (1953) study, with the findings from the present study, may help to provide a partial answer to this question. He required his Ss to perform sample tasks representing each factor extracted and then to rate their use of imagery during the performance of these tasks. It was found that high-imagers performed better than low-imagers on the Spatial Manipulations tasks but the performance of the two groups on the Spatial Recognition and Spatial Reasoning tasks was not differentiated. Barratt indicates that these results justify the use of these tests for measuring imagery. In view of the cross-validation achieved by differences in performance of the two imagery groups on the tasks representing the separate factors his reasoning is warranted, at least in part. Nevertheless, in view of the present findings regarding the possibility that introspective reports are partially confounded with social desirability, complete answers to the validity of the imagery construct can only come from further development of a carefully constructed nomological net, a part of which must necessarily be comprised of the findings from Barratt's and from Stewart's studies.

The separate extraction of the Automatization factor from the imagery factor, was in a sense, predictable from Broverman's (1964) studies. On the surface it may appear that both the Gottschaldt Figures Test and the Stroop Color-Word Test are similar to the extent that they involve interference of performance by the presence of irrelevant stimuli. However, the distinctiveness of the tasks lies in the relationship between the interfering stimuli to the critical stimuli in each. Accordingly, in the Stroop Color-Word Test the stimulus attribute is readily perceived. Successful performance is dependent on a set to respond to certain obvious features of the stimulus and not to respond to the wrong, interfering, though equally obvious features. This set-to-respond in certain ways is undoubtedly present in Automatization. It is clearly distinct from the reorganization of stimulus structure required in Imagery, as represented in the Gottschaldt test.

The Verbal factor identified in both studies is so familiar that very little additional explanation to that already provided in textbooks and manuals seems necessary. As one can readily see, it is the factor comprised of acquired verbalizations and language symbols and the ability to employ these symbols, in various ways, within tasks where general mental ability and symbolic transformations of incoming stimuli facilitate performance. Similarly, the extraction of the Anxiety factor in Study I was not unanticipated since, in two ways, it was quite unlike any of the other measures. First, it was a self-report or introspective measure rather than an objective test and second, the questions related to the affective domain of behaviors rather than to the cognitive domain. Nevertheless, the extraction of the Verbal and Anxiety factors are important to the present discussion to the extent that, except for the loading of the SAT:Math score on Imagery, none of the other loadings on these factors overlapped significantly with those of Imagery. Thus, the data provide further evidence that Imagery is a separate constituent of cognitive structure or a separate cognitive strategy from Verbal ability.

The final factor to be discussed is Social Desirability. It was not altogether surprising that one component of this factor, the introspective reports of ability to control imagery and vividness of imagery loaded heavily on a factor other than that represented by the objective tests of imagery. However, it was surprising that they should be heavily weighted on Social Desirability. Inspection of the items on these scales suggest a possible explanation: The scales for the introspective reports of imagery may imply to some Ss that to be able "to control one's imagery" and "to experience vivid images" is a highly valued characteristic comparable, for example, to possessing a high I.Q. The Social Desirability scale also measures qualities of behavior that reflect dependence on the acceptance, recognition and approval of others. If ability to conjure up images is believed to be a culturally desirable trait then it is consistent that this bias will affect the scores on the scale. It is most interesting to note that Richardson (1969, p. 87) does indicate a correlation between the richness of fantasy (measured by introspective reports) and persuasibility (defined as readiness to accept social influence regardless of what is known about the communicator). However, nowhere in the book could the present investigator find where response bias, need for approval, or similar behavioral qualities were considered in the interpretation of data presented. Nevertheless, it was clear that differential performances such as "perceptual achievements, ... involved in responding to an ambiguous ink blot or in recognition of an object" (Richardson, 1969, p. 131) and other similar behaviors attributed by Richardson to differences between high and low visualizers could also be influenced by social desirability and thereby would provide alternative explanations.

In summary, the results of the present investigation appear to warrant two conclusions. First, imagery as defined by objective tests appears to be a distinctive individual difference variable. Relating underlying processes such as those tentatively described above to performance in situations predominantly concrete or abstract should be a fruitful source of hypotheses for further investigations. In this context, the present investigator wishes to reemphasize the caution certainly familiar to potential investigators, that is, hypotheses should be based on processes assumed to underlie the factors rather than on the labels attached to the factors. At best, such labels reflect the idiosyncracies of the investigator and by themselves are more often misleading than not. A case in point is the ambiguity associated with a term such as automatization which can be interpreted variously as "readiness to respond," "rigidity," "set to respond to given attributes," or even perhaps as "fluency in translating pictures into words." We have already pointed out in the introduction the numerous definitions attached to imagery. The term Imagery, even as it is used here, does no more than imply a non-linguistic category. The question of whether it is an ikon, engram, or non-linguistic meaning category is certainly unresolved. Our present inclin ation is to restrict its definition to those processes presumably involved in test performance. Still further refinement of these definitions appear to be imperative if the elusive aptitude by treatment interactions are to be captured in systematic investigations.

A second conclusion from the present study is that introspective reports, as measures of imagery, do not possess construct validity. This conclusion was implied not only by the results of the present investigation but by the examination of data presented in reports of studies where introspective scales were employed. Where such introspective scales have been employed, as they were in studies reported by Richardson (1969), the data crediting differences in performance to differences in imagery should be interpreted with caution. Some consideration may be given to further development of these scales with an attempt to remove response bias or their heavy dependence on social desirability. For the present, they must be considered to be confounded with response bias.

This study provides only a description of one structure of imagery. It does not answer important and interesting developmental questions such as the degree of imagery relative to verbal ability in children compared to adults. Nor does it indicate how much the use of imagery reflects a skill as it probably does with adults. These kinds of distinctions suggest interaction with manipulations of task characteristics (such as concreteness) in the former and with manipulations of motivational levels in the latter. The interpretations provided here should provide a basis for further factorial studies to differentiate among imagery structures or specific kinds of imagery that may vary for the senses. However, the primary interest is to determine whether imagery as isolated here is descriptive of intellectual performance that transcends purely perceptual effects. In view of the potential fruitfulness of the approach and the current popularity of imagery as a cognitive process, investigations of the antecedent conditions associated with it and of its interactions with task and stimulus variables are in order.

References

Alpert, R., & Haber, R. N. Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology*, 1960, *61*, 207-215.

Barratt, P. E. Imagery and thinking. *Australian Journal of Psychology*, 1953, *5*, 164-174.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. *Differential aptitude tests*. (Grades 8-13 and adults.) New York: Psychological Corporation, 1963.

Betts, G. H. *The distributions and functions of mental imagery*. New York: Teachers College, Columbia University, 1909.

Bower, G. H. Analysis of a mnemonic device. *American Scientist*, 1970, *58*, 496-510.

Broverman, D. M. Generality and behavioral correlates of cognitive styles. *Journal of Consulting Psychology*, 1964, *28*, 487-500.

Broverman, D. M., Kleiber, E. L., Kobayashi, Y., & Vogel, W. Roles of activation and inhibition in sex differences in cognitive abilities. *Psychological Review*, 1968, *75*, 23-50.

Budner, S. Tolerance of ambiguity as a personality variable. *Journal of Personality*, 1963, *30*, 29-50.

College Entrance Examination Board Scholastic Aptitude Test. Princeton, New Jersey: Educational Testing Service, 1962-1963.

Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist*, 1957, *12*, 671-694.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychological Bulletin*, 1955, *52*, 281-302.

Crowne, D. P., & Marlowe, D. *The approval-motive: Studies in evaluative dependence*. New York: Wiley, 1964.

Galton, F. Statistics of mental imagery. _Mind_, 1880, _5_, 301-318.

Galton, F. _Inquiries into human faculty and its development_. London: J.M. Dent & Sons, 1883. Reprinted New York: Everyman, 1911.

Gordon, R. An investigation into some of the factors that varies the formation of stereotyped images. _British Journal of Psychology_, 1949, _39_, 156-167.

Graham, F. K., & Kendall, B. S. Memory-for-Designs Test: Revised General Manual. _Perceptual and Motor Skills_, 1960, _11_, 147-188.

Hollenberg, C. K. Functions of visual imagery in the learning and concept formation of children. _Child Development_, 1970, _41_, 1003-1015.

Holt, R. R. Imagery: the return of the ostracized. _American Psychologist_, 1964, _19_, 254-264.

Jensen, A. R., & Rohwer, W. D., Jr. The Stroop color-word test: A review. _Acta Psychologica_, 1966, _25_, 36-93.

Kaiser, H. F. The Varimax criterion for analytic rotation in factor analyses. _Psychometrika_, 1958, _23_, 187-200.

Lamke, T., Nelson, M. J., & Kelson, P. C. _The Henmon-Nelson tests of mental ability. Revised edition_. Boston: Houghton-Mifflin, 1931-1960.

Lindsay, C. A., Williams, G. D., & Peterman, D. J. Technical manual for the 1969 _Division of Counseling_ pre-registration testing program. University Park, Penna.: Pennsylvania State University, 1969. Mimeo.

Mednick, S. A., & Mednick, M. T. _Remote associates test_. Boston: Houghton-Mifflin, 1967.

Paivio, A. On the functional significance of imagery. _Psychological Bulletin_, 1970, _6_, 385-392.

Richardson, A. _Mental imagery_. New York: Springer, 1969.

Rohwer, W. D., Jr. Images and pictures in children's learning: Research results and educational implications. _Psychological Bulletin_, 1970, _6_, 393-403.

Rokeach, M. The open- and closed-mind. New York: Basic Books, 1960.

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, University of Toronto, 1965.

Stroop, J. R. Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 1935, 18, 646-660.

Thurstone, L. L. A factorial study of perception. Chicago: University of Chicago Press, 1944.

Thurstone, L. L., & Jeffrey, T. G. Space thinking (Flags). Chicago: Education-Industry Service, 1959.

Wechsler, D. Wechsler Adult Intelligence Scale. New York: Psychological Corporation, 1955.

Imagery Ability, Abstractness and Word Order
As Variables in Recall of Adjectives and Nouns

Francis J. Di Vesta and Steven M. Ross

Among the empirical results of experiments on item imagery-arousal are the findings that word pairs are more easily learned in the noun-adjective order as opposed to adjective noun order (Lambert & Paivio, 1956; Paivio, 1963; Kusyszyn & Paivio, 1966; Yuille, Paivio & Lambert, 1969); that concrete nouns, but not abstract nouns are more effective associative cues than adjectives (Lockhart, 1969); and that pictures are easier to remember than concrete nouns (Stewart, 1965; Paivio, 1969). Although investigations of imagery as a stimulus attribute has yielded generally consistent and reasonably conclusive evidence, there is considerably less empirical support for theoretical notions regarding imagery as a transformational process influenced by individual differences in imagery-ability.

A potentially fruitful approach in defining imagery abilities has involved the use of spatial manipulation tasks. Using these measures, Ernest and Paivio (1969; 1970) found that female high imagers are more accurate in recalling incidental material than female low imagers, and that high imagers of both sexes have greater reaction times in associating abstract stimuli than in associating concrete stimuli. Stewart (1965) too, has reported that the performance of high imagers was facilitated by concrete stimulus materials in several learning situations.

The present investigation was, in part, a replication of one by Yuille, Paivio, and Lambert (1969) in which the order of presenting paired-associates

PRECEDING PAGE BLANK

was manipulated; that is, Ss were presented pairs of words in either the noun-adjective or adjective-noun order. In addition, controls were provided for the abstract-concrete characteristics of the adjectives as well as of the nouns. It was found that recall was superior for the noun-adjective order, for high-Imagery rather than low-Imagery stimulus elements, and for high-Imagery rather than low-Imagery response elements. Overall, stimulus Imagery emerges as the most critical factor in accounting for differences in item recall.

The main purpose of the present study, however, was to extend the one conducted by Yuille, et al., (1969) by incorporating levels of imagery, as an individual difference variable. This factor was included in order to investigate the hypothesis that levels of imagery ability interact with treatments to affect performance. Specifically, it was hypothesized that, relative to low imaging ability, high imagery ability would facilitate the learning of abstract stimuli more than of concrete stimuli. These hypotheses were suggested in several studies by Paivio and his colleagues. For example, Paivio and Foth (1970) demonstrated that mediation instructions emphasizing imaginal processes resulted in greater recall of abstract pairs than mediation instructions emphasizing verbal processes. Ernest and Paivio (1971) showed that the relative superiority of high imagers over low imagers in reaction speed was greater when the stimuli to be associated were abstract for both imaginal and verbal instructions. In general, these findings suggest that imaginal processes complement verbal associations with the presentation of abstract stimuli. Concrete stimuli, on the other hand, elicit easily detectable cues that are as accessible for dual processing along with verbal cues by high imagers and non-imagers alike. However, non-imagers are presumably less able to establish pictorial associations and, as a result, are less efficient at retrieving the particular response when abstract stimuli are used. Therefore it was hypothesized that high imagers would have relatively greater recall for abstract pairs than would low imagers.

Method

Design

The Ss were presented two study-recall trials of a list of paired-associates. The word pairs appearing in a given list were presented in either a noun-adjective (N-A) or adjective-noun (A-N) order. Thus, in one condition, nouns served as the stimulus elements and adjectives as the response elements. In the other condition, adjectives served as stimuli while nouns served as response elements. Imagery arousal (I) of the noun associates was employed as a within-subjects variable with two levels. Using the rating scales by Paivio, Yuille, and Madigan (1968), half of the nouns were selected on the basis of high imagery and concreteness, and the other half were selected on the basis of low imagery and abstractness. This variable was orthogonally crossed with a similar set of conditions in which the imagery arousal (I) of the adjective associates as another within-subjects variable was manipulated. Thus, within each order (N-A or A-N), a given list of word-pairs was comprised of equal numbers of word pairs in which both members of a pair had high rated imagery, in which one member had high rated imagery and the other had low rated imagery, and in which both members had low imagery. The between-subjects variable of list order was orthogonally crossed with another between-subjects variable, that of imagery ability as an individual difference variable. Thus, half of the Ss were classified as high imagers and the other half as low imagers according to their performance on spatial relations tests. All Ss were given two study-recall trials. In summary, these manipulations implied a 2 x 2 x 2 x 2 x 2 analysis of variance with the between-subjects factors being two orders of presentation (N-A and A-N), and two imagery aptitude groups (high-imagers and low-imagers); and the within-subjects factors being two levels of stimulus imagery (High-I and Low-I), two levels of response imagery (High-I and Low-I), and two study-recall trials.

Subjects

The S pool for this experiment consisted of approximately 300 undergraduates enrolled in an introductory educational psychology course at the Pennsylvania State University. Sixty-five high imagers and 65 low imagers were selected as potential Ss according to their average standard score (T) on a test battery of three spatial relations tests: Flags: A Test of Spatial Thinking (Thurstone & Jeffrey, 1956), A Space Relations Test from the Differential Aptitude Test (Bennett, Seashore, & Wesman, 1947), and the Gottschaldt Figures Test as described by Thurstone (1944). Of these, 54 high imagers (average T-score = 59.50) and 54 low imagers (average T-score = 37.75) agreed to participate in the present experiment, and were extended credit towards the course grade. The experimental variations were administered to Ss in groups of 3. All Ss were randomly assigned to separate conditions upon entry into the laboratory.

Word Lists

A stimulus list consisted of 24 adjective-noun (A-N) pairs; six pairs of which were comprised of high-imagery adjectives and high-imagery nouns (HH); six pairs were high-imagery adjectives and low-imagery nouns (HL); six pairs were low-imagery adjectives and high-imagery nouns (LH); and six pairs were low-imagery adjectives and low-imagery nouns (LL). Three randomized lists were prepared, thus requiring a pool of 36 adjectives and 36 nouns. Items for the pairs were randomly selected from this pool without replacement. The noun-adjective (N-A) lists were formed by reversing the pairs in the A-N lists. Altogether six lists were prepared, each with two random orders of presentation. No word-pair occurred more than once in either the A-N or N-A Presentation Order.

The 18 high-imagery (concrete) nouns had average ratings of 6.40 on the imagery scale and 6.80 on the concreteness scale; the 18 low imagery (abstract) nouns had imagery and concreteness ratings of 3.50 and 1.77 respectively. These words were selected from the Paivio, Yuille, and Madigan (1968) norms. Separate norms for a set of 75 adjectives and five nouns were obtained specifically for this study inasmuch as ratings of adjectives on the imagery and concreteness scale were not readily available. The five nouns were selected at random from the Paivio et al. (1968) norms in order to provide a basis for determining similarity in ratings for the two groups. The set of 75 adjectives and five nouns were rated for imagery arousal by 15 undergraduate volunteers who were from the same S pool as that employed for the experiment. On the basis of their ratings, 18 High-I adjectives averaging 6.29 on a 7-point scale, and 18 Low-I adjectives averaging 2.74 were selected. The average ratings of the five nouns was 4.22 which did not differ significantly from the 4.02 obtained by averaging the corresponding I scores from the Paivio et al. (1968) norms. Thus, it was assumed that the ratings of adjectives were comparable to the ratings of nouns available from Paivio, et al.'s norms.

Following the selection of the individual words, each adjective was paired with both a High-I and a Low-I noun. These pairs are displayed in Table 1. Each word-pair was printed on a 3 x 5 in. card.

Table 1

Nouns and Adjectives used in the Paired-Associate Lists

Adjectives*		Nouns	
High Imagery	Low Imagery	High Imagery	Low Imagery
Dark	Boring	Elephant	Satire
Hairy	Known	Fox	Spirit
Round	Bad	Camp	Edition
Bumpy	Subtle	Hurdle	Crisis
Vertical	Mature	Acrobat	Quality
Rocky	Usual	Volcano	Disposition
Fuzzy	Trite	Clothing	Jealousy
Blue	Hungry	Reptile	Obsession
Sharp	Best	Corner	Anger
Colorful	Actual	Factory	Honor
Cloudy	Personal	Alcohol	Belief
Glassy	Different	Skillet	Pleasure
Smooth	Populer	Jelly	Intellect
Shiny	Quiet	Revolver	Memory
Wet	Obvious	Barrel	Sensation
Burnt	Dominant	Headlight	Betrayal
Small	Tardy	Whale	Idea
Bloody	Real	Tweezers	Virtue

* Each adjective (High-I and Low-I) was paired with the corresponding High-I and Low-I nouns.

Procedure

Three Ss were scheduled for each experimental session and were seated at separate locations in the laboratory. Prior to Ss arrival at the session, one of the six decks of 3 x 5 inch cards had been placed randomly at each of the three positions. Thus, each S within a group received a unique deck consisting of two study-recall trials in either the A-N or N-A order.

All Ss were given standard-paired associate instructions administered by means of a tape recorder. The E was always present to supervise the general administration and to respond to questions regarding the procedure. The basic task was identical for all treatment conditions. At the sound of a bell, S viewed the first pair by flipping over the top index card. Ten seconds were allotted to study each pair after which S was again directed by the bell to proceed to the next card. Verbal instructions announced the end of the study list and the beginning of the recall session. During recall only the stimulus elements of the individual word pairs appeared on the index cards. An interval of 15 seconds was provided for S to write down the appropriate response element. Cover sheets were used to obscure all previous answers. When the recall trial was completed the answer sheet was collected and E proceeded to administer the second study-recall presentation of the same pairs via the same procedure employed for the first trial.

Results

The number of correct responses were analyzed by a mixed analysis of variance with two between-subjects and three within-subjects variables. The between-subjects factors were two Presentation Orders (N-A and A-N), and two levels of Imagery (high imagers and low imagers); the within-subjects factors were the two levels of Stimulus Imagery (High I and Low I) two levels of Response Imagery (High I and Low I), and two Recall Trials.

The significant main effects were as follows: The effect due to Presentation Orders yielded $\underline{F}$ (1,104) = 15.09, $\underline{p} < .001$, indicating that the N-A order ($\bar{X}$ = 4.36) was superior to the A-N order ($\bar{X}$ = 3.58) across conditions. The effect due to Stimulus Imagery yielded $\underline{F}$ (1,104) = 134.96, $\underline{p} < .001$. The effect due to Response Imagery yielded $\underline{F}$ (1,104) = 36.17, $\underline{p} < .001$. These findings imply that High-I words were more easily associated than Low-I words whether in stimulus or response positions. As would be expected the effect due to Recall Trials was highly significant yielding $\underline{F}$ (1,104) = 532.35, $\underline{p} < .001$. These main effects and the non-significant ($\underline{p} > .05$) effect associated with Imagery levels were qualified by the interactions discussed in the paragraphs that follow.

The effect due to the interaction between Presentation Order and Recall Trials yielded $\underline{F}$ (1,104) = 6.09, $\underline{p} < .05$. Though more words were correctly recalled during the second recall trial for both the N-A and A-N orders, the degree of improvement was more pronounced for the A-N condition. This effect is almost certainly attributable to the near ceiling performance of $\underline{S}$s in the N-A condition during the first recall trial ($\bar{X}$ = 4.68). If this is contrasted with the average recall in the A-N condition ($\bar{X}$ = 2.48), it is obvious that comparative potentialities for improvement were markedly uneven (the ceiling score was 6.00). The same interpretation can probably be applied to the effect due to the interaction between Stimulus Imagery and Recall Trials which yielded $\underline{F}$ (1,104) = 11.98, $\underline{p} < .001$; the high-I stimulus pairs were recalled at a near ceiling level during the first recall trial. Nevertheless, these findings indicate that the effect of imagery is readily demonstrable during the initial stages of learning.

The interaction between Presentation Order and Response Imagery yielded $\underline{F}$ (1,104) = 5.05, $\underline{p} < .05$. This finding indicates that, in the response position, noun imagery is a more critical factor than adjective imagery. In contrast, the effect due to the interaction between Presentation Order and Stimulus Imagery was not significant ($p > .05$). Thus, imagery is a more important variable in the stimulus than in the response element of a word pair, whether the stimulus is a noun or adjective. This result shown in Figure 1 corroborates the findings of Paivio and his associates (Yuille, Paivio, & Lambert, 1969; Paivio, 1970).

Further support for the "stimulus peg" notion is provided by analyses of the effect due to the first order interaction between Stimulus Imagery and Response Imagery which yielded $F\ (1,104) = 20.25$, $p < .001$. The order of difficulty for learning under the various stimulus-response imagery conditions (from easiest to most difficult) was High-I Stimulus - High-I Response ($\overline{X} = 4.95$), High-I Stimulus - Low-I Response ($\overline{X} = 4.25$), Low-I Stimulus - High-I Response ($\overline{X} = 3.38$), and Low-I Stimulus - Low-I Response ($\overline{X} = 3.29$). These results indicate that increasingly greater gains were demonstrated as the pairs increased in concreteness. Thus, a High-I response element had a greater facilitative effect when the stimulus element was also a High-I noun or adjective. However, the significant triple interaction between the above factors and recall trials provides further qualification of this conclusion. This interaction, graphically displayed in Figure 2, is mostly attributable to the previously discussed ceiling effect that occurred in Trial 2. However, another differential effect between trials occurred in the ordering of the means where the

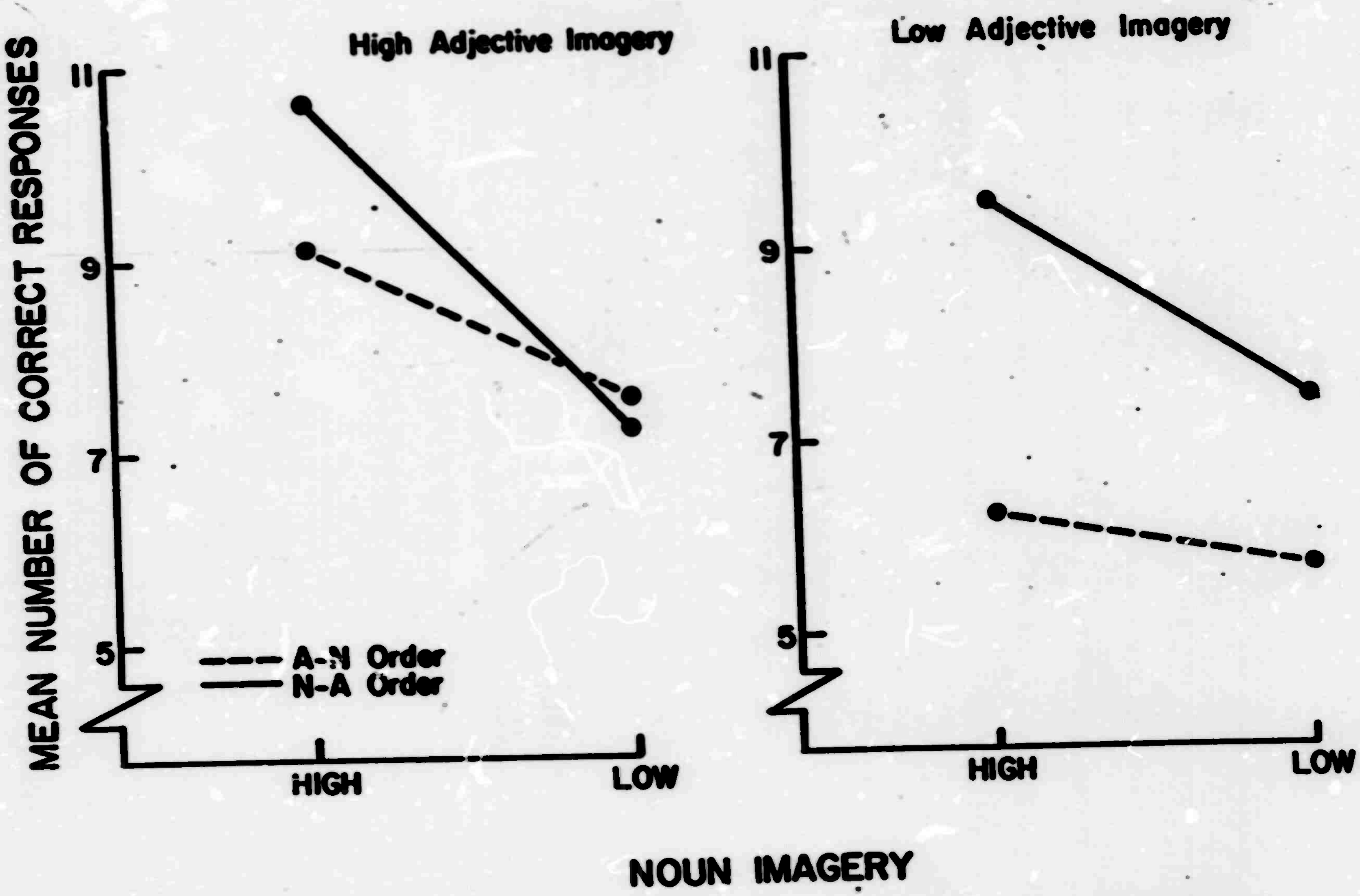

Figure 1. Mean number of correct responses, for trials 1 and 2 combined, as a function of noun imagery, adjective imagery, and word order.

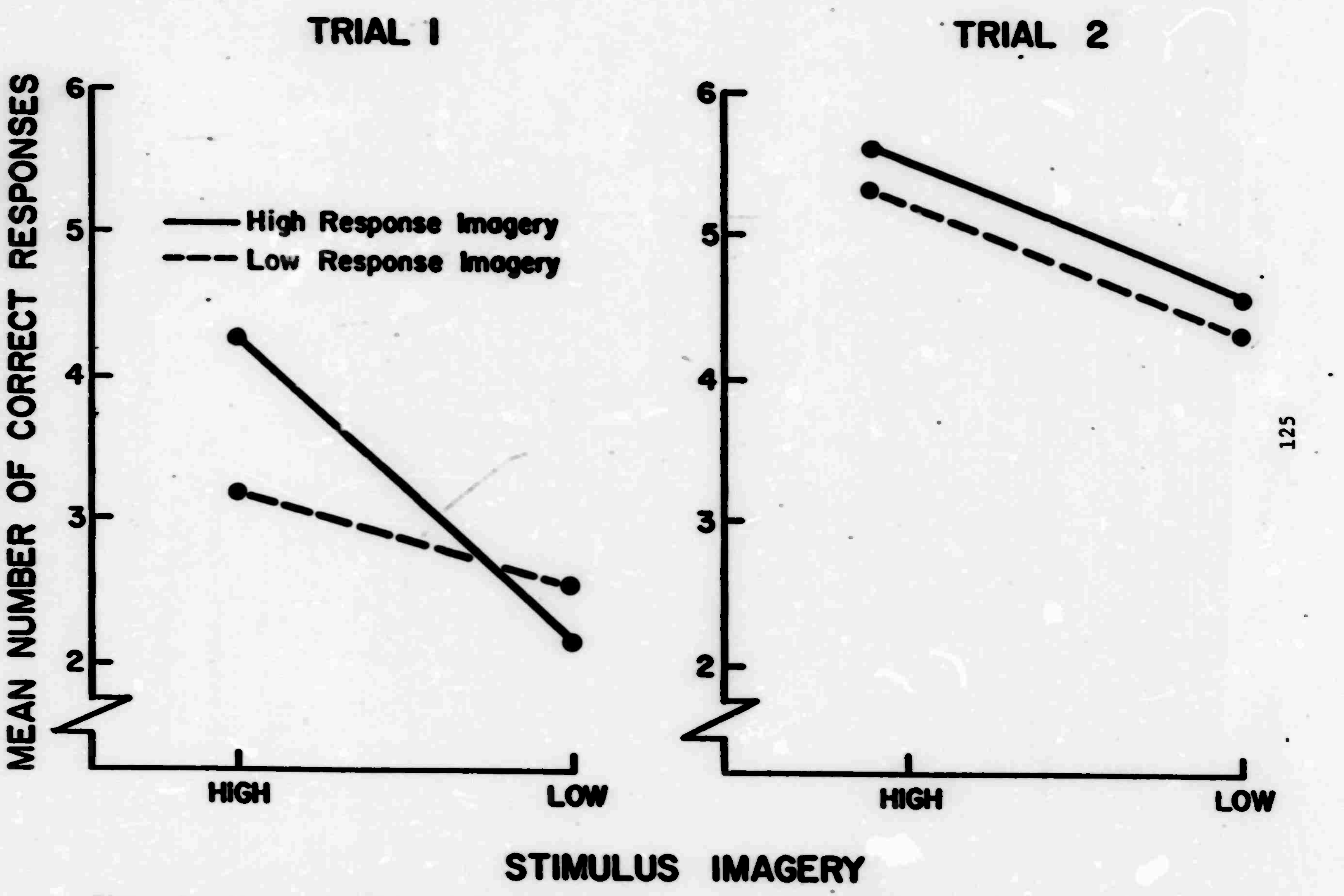

Figure 2. Mean number of correct responses for trials 1 and 2, as a function of stimulus imagery and response imagery.

hypothesized superiority of Low-I - High-I pairs over Low-I - Low-I pairs was reversed in trial 1, although they were not significantly different ($t < 1.00$). A subsequent analysis was performed on the data in the cells described above to test directly the hypothesis that imagery is more effective on the stimulus side than on the response side of paired-associates. In effect, all pairs with High-I stimulus elements ($\overline{X} = 3.73$ in Trial 1 and $\overline{X} = 5.48$) in Trial 2) were compared with those containing Low-I stimuli ($\overline{X} = 2.37$ in Trial 1 and $\overline{X} = 4.45$ in Trial 2), and similarly for the High-I ($\overline{X} = 3.22$ in Trial 1 and $\overline{X} = 5.12$ in Trial 2) and Low-I ($\overline{X} = 2.88$ in Trial 1 and $\overline{X} = 4.81$ in Trial 2) response elements. The overall comparison indicated that in both Recall Trial 1 ($t = 9.62$, $p < .001$) and Recall Trial 2 ($t = 6.57$, $p < .001$) stimulus imagery was a more critical determinent of recall facility than was response imagery .

Imagery aptitude was involved in a significant four-way interaction with Presentation Order, Stimulus Imagery, and Recall Trials [F (1,104) = 4.75, $p < .05$]. A summary of the means related to this interaction is presented in Table 2. Although the effects are not extreme, it is apparent that High Imagery ability was most effective in the recall of Low-I stimulus pairs. This result was more apparent in the A-N order during Trial 1 and in the N-A order during Trial 2. It is difficult to account for this order reversal between trials, but it is probably of questionable theoretical significance when the ceiling effect described above is taken into consideration. It should be noted that the performance of Low Imagers surpasses that of High

Table 2

Mean Number of Correct Responses on Trial 1 and Trial 2

As a Function of Imagery Aptitude, Presentation Order, and Stimulus Imagery

Stimulus Imagery and Imagery Aptitude	Presentation Order			
	Adjective-Noun		Noun-Adjective	
	Trial 1	Trial 2	Trial 1	Trial 2
Concrete Stimulus				
High-Imagers	3.02	5.37	4.56	5.81
Low-Imagers	3.30	5.00	4.07	5.72
Abstract Stimulus				
High-Imagers	2.02	4.37	2.80	4.98
Low-Imagers	1.57	3.98	2.46	4.48

Imagers when the High I stimulus was an adjective (A-N order). Thus, the data directionally support the hypothesis that the superiority of High Imagers over Low Imagers is greater when the stimulus is Low I. Additional support for this hypothesis was provided by the interaction between Imagery, Stimulus Imagery, Response Imagery and Recall Trials which yielded F (1,104) = 2.62, $p > .10 < .20$. Though this interaction was not significant, and indicates only a tendency, it suggests that the superiority of High Imagers over Low Imagers was most pronounced for Low-I - High-I pairs in Trial 1 (mean difference ($\bar{d}$) = +.58), and for Low-I - Low-I pairs in Trial 2 ($\bar{d}$ = +.51). On the other hand, the smallest differences between the two groups appeared in the learning of High-I - High-I pairs in Trial 1 ($\bar{d}$ = -.02) and also in Trial 2 ($\bar{d}$ = +.18). None of the other main effects or interactions were found to be significant.[1]

Discussion

The results of the present study provide clear support of Yuille, et. al's (1969) findings that imagery of nouns influences paired-associate learning more than does imagery of adjectives. With high imagery adjectives, nouns in the N-A order were more influential than were nouns in the A-N order. However, it is interesting that low imagery nouns coupled with high imagery adjectives were more effective in facilitating performance in the A-N than in the N-A order. When low imagery adjectives were employed, the N-A order was more effective than the A-N order whether high or low imagery nouns were used. These results imply that it is the rated imagery of the stimulus member of the pair that is most influential in its effect on rate of learning and retention. It is the concreteness of the stimulus rather than its form class that is the important variable. In contrast, the effect of meaningfulness of the response member has been found to be more influential in facilitating performance (Underwood & Schulz, 1960).

Although the results concerning individual differences in imagery ability were only suggestive, they are of theoretical importance. As Ernest and Paivio (1969) indicate, "Whether the imagery hypothesis may be extended to encompass individual differences in imagery ability is of considerable theoretical import. Successful predictions of performance based on such differences would provide further convergent evidence that a common intervening process is involved whether imagery is defined by stimulus attributes, mnemonic instructions, or individual differences" (p. 181). The present data indicate that high-imagers have an advantage over low-imagers when the stimulus is of low rated imagery. On the other hand, there is less difference in performance between the two groups when the stimulus is of high rated imagery. From these data, it appears that imagery ability affects performance for the same reasons that concrete stimuli do. When stimuli are concrete there is no further advantage to be gained by having high imagery ability; the stimuli are equally discriminable to both high and low imagers. However, when stimuli are of low-rated imagery, that is, when they are abstract, their ambiguity can be lessened by the imaginal ability of the high imager.

Though these data by themselves are of marginal significance, they gain importance when coupled with the results of earlier studies. Thus, Ernest and Paivio (1969) found that incidental recall was consistently better for high imagers than for low imagers. They (Ernest & Paivio, 1971) also found that, as measured by reaction times to elicit a verbal associate or to arouse an image, the high imager's performance was superior (i.e., latencies were shorter) to that of the low imager when the stimuli were abstract. The findings of the present study support this result. Of further support to this hypothesis is the study by Paivio & Foth (1970) whose <u>Ss</u> were <u>required</u> to either write sentences or draw pictures for a verbal mediation condition, or an imagery mediation condition, respectively. They found that verbal mediation facilitated the learning of abstract pairs, while imagery facilitated the learning of concrete pairs. However, in another condition, the _Ss_ were merely provided mediation instructions (i.e., to generate either verbal or imaginal mediators) but were not _required_ to employ _sentences_ or _drawings_. Under this circumstance imagery was found to be better than verbal mediation for abstract pairs. This finding suggests that abstract pairs can be more easily learned with the aid of imagery processes. Thus, if imagery instructions are available but not forced upon _S_ (Paivio & Foth, 1970) or if _S_ has high imagery ability (Ernest & Paivio, 1971) there will be a positive effect upon _abstract pair_ learning. Other evidence suggests that imagery ability may affect learning of abstract pairs differently from the recall of pictorial or verbal stimuli. Thus, Stewart (1965) and Kuhlman (1960) found that high-imagers recalled more items presented in pictorial than in verbal form, while the low-imagers recalled more verbal than pictorial items. Accordingly, it is hypothesized that acquisition and retrieval strategies generated in the free-recall task employed by Stewart and Kuhlman differ from those generated in a word-association or paired-associate learning task. Clustering and subjective organization based on pictures versus words in a free-recall task implies a preference for a given strategy (i.e., imaging) over another (e.g., employing verbal mediators). On the other hand, the effect of imagery in paired-associate learning implies the ability of the individual to employ a strategy that efficiently transforms the stimulus to a form necessary for effective hook-up with the response. Although imagery ability appears to be functionally related to learning and memory there is still insufficient evidence to indicate that this ability reflects the same process as that suggested by experimental manipulations of word imagery (Ernest and Paivio, 1969, p. 182). Nevertheless, the data from the present experiment strongly suggest that this may be the case.

Footnotes

[1] In a previous analysis Presentation List was included as a factor. It was not significant as a main effect but was involved in a significant interaction with Stimulus Imagery ($p < .05 > .01$) and in a five way interaction with Presentation Order, Individual Aptitude, Recall Trials, and Response Imagery ($p < .05 > .01$). Because the items in the list were selected at random there was no explanation for these differences. Inasmuch as the differences were not disordinal (in the interaction with Stimulus Imagery) or systematic in the five-way interaction, these interactions were disregarded in subsequent analyses.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential Aptitude Tests. New York: The Psychological Corporation, 1947.

Ernest, C. H., & Paivio, A. Imagery ability in paired-associate and incidental learning. Psychonomic Science, 1969, 15, 181-182.

Ernest, C. H., & Paivio, A. Imagery and sex differences in incidental recall (Research Bulletin No. 121) London Ontario: Department of Psychology, University of Western Ontario, 1970.

Ernest, C. H., & Paivio, A. Imagery and verbal associative latencies as a function of imagery ability. Canadian Journal of Psychology, 1971, 17, 83-90.

Kuhlman, C. K. Visual imagery in children. Radcliffe College, Unpublished doctoral thesis, 1960.

Kusyszyn, I., & Paivio, A. Transition probability, word order, and noun abstractness in the learning of adjective-noun paired-associates. Journal of Experimental Psychology, 1966, 71, 800-805.

Lambert, W. E., & Paivio, A. The influence of noun-adjective order on learning. Canadian Journal of Psychology, 1956, 10, 9-12.

Lockhart, R. S. Retrieval asymmetry in the recall of adjectives and nouns. Journal of Experimental Psychology, 1969, 79, 12-17.

Paivio, A. Learning of adjective-noun paired-associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370-379.

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Paivio, A. On the functional significance of imagery. Psychological Bulletin, 1970, 6, 385-392.

Paivio, A., & Foth, D. Imaginal and verbal mediators and noun concreteness in paired-associate learning: the elusive interaction. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 384-390.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology, 1968, 76, (No. 1, Pt. 2).

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, University of Toronto, 1965.

Thurstone, L. L. A factorial study of perception. Chicago: University of Chicago Press, 1944.

Thurstone, L. L., & Jeffrey, T. E. Flags: A Test of Spatial Thinking. Chicago: Industrial Relations Center, 1956.

Underwood, B. J., & Schulz, R. W. Meaningfulness and verbal learning. New York: Lippincott, 1960.

Yuille, J. C., Paivio, A., & Lambert, W. E. Noun and adjective imagery and order in paired-associate learning by French and English subjects. Canadian Journal of Psychology, 1969, 23, 459-466.

The Retrieval of Abstract and Concrete Materials By High and Low Imagers Employing Imaginal and Verbal Mediation With Abstract and Concrete Mnemonic Aids

Francis J. Di Vesta and Phyllis M. Sunshine

On the basis of a review of an extensive number of studies, Paivio (1969b) has suggested that a two stage model was implied for associative learning of noun pairs. Thus, the meaning of concrete nouns is acquired through both direct experience with the referent and association with other words. The consequence is that either verbal associations or nonverbal images might be evoked by these nouns to influence learning and recall. The meanings of abstract words, on the other hand, are acquired primarily, if not exclusively, through intraverbal experience. Accordingly, abstract words tend to elicit verbal rather than imaginal associations. Stated in a slightly different manner, the learner confronted with a task consisting of associating concrete nouns might employ both verbal and imaginal processes or strategies. However, either because he has a preference for using images or because images are more available than words, the learner tends to use the imagery strategy with concrete words. This conclusion is supported by both the subjects' subjective reports and by comparison of learning scores under the two strategies. Confronted with a task in which he is required to associate abstract terms, the learner employs the strategy emphasizing verbal association.

While a number of experimental procedures have been used to test these hypotheses, two methods in particular are of concern in the present study. One of these methods employs the "one is a bun" mnemonic device for facilitating recall. The jingle in this mnemonic aid can be composed of concrete pegwords, in which case it should facilitate the use of an imagery strategy, or it can be composed of abstract pegwords thereby facilitating the use of a verbal strategy. In the other research method of concern here instructions are employed to induce the subject to use either imagery or verbal associations when associating the word-pairs. On occasion both repetition-set and no-set conditions have been incorporated into the design as controls (e.g., Paivio & Yuille, 1969). Inasmuch as comparison with the latter two treatments indicate a clear superiority of verbal and imaginal processes for learning abstract and concrete pairs, respectively, the control conditions were not considered further for the present study.

Induction of these sets, through instructions only, does not always provide a strong effect. Accordingly, the procedure had been modified by Paivio & Foth (1970) by requiring the subject, in the imaginal set treatment, to draw a picture linking the pegword to its serial counterpart in the list to be learned; and by requiring the subject, in the verbal set treatment, to write a sentence using the two words. The importance of these methodological variations for the present investigation is that a means by which strategies can be manipulated is provided. Since both procedures influence the use of strategies in the same way (which, incidentally, is comparable to the effect of concreteness-abstractness described above) it was hypothesized that the use of the two methods in a crossed design, where the subject's task was to learn concrete and abstract lists, would result in further enhancement of the effect.

By employing these two treatments in conjunction with groups of high and low imagers it was hypothesized that the strategies used by subjects would become explicit in the ability of the two groups to learn and recall concrete and abstract lists. More specifically, it was hypothesized that since both verbal and imaginal processes could be used effectively with concrete materials, both groups of subjects would perform equally well on the task (see Ernest & Paivio, 1971). However, since the abstract list was assumed to be more easily associated to other words by verbal processes, the performance of the low-imagers would be especially hindered when forced to employ imaginal processes with the abstract mnemonic in learning an abstract list. Thus, the present study was an attempt to explore the possibility that "... the [effects of the] three classes of independent variables [in studies of imagery] - stimulus attributes, experimentally manipulated mediators, and individual differences [in imagery-ability]-... are mediated by common intervening processes" (Paivio, 1969b, p. 259).

Method

Design

The overall design of this study required that the subjects first learn (memorize) a jingle which was to serve as a conceptual peg for later learning tasks. Half of the subjects learned a jingle in which the critical words were concrete; the remaining subjects learned a jingle in which the critical words were abstract. These two treatments were orthogonally crossed with two mediational modes. Thus, when using the jingle as a mnemonic aid half of the subjects were required to employ it within an imagery-set, that is, they were to literally draw pictures, while the other half were to employ the jingle in a verbal set, that is, they were to write sentences to link the elements of the jingle with the new material to be learned. The jingles and sets were used to learn four different lists of words, two of which were comprised of concrete nouns and the other two were comprised of abstract nouns. Tests for recall of each list were administered immediately after learning the list and tests for recall of words in all lists were administered at the conclusion of the experiment. The subjects were selected on the basis of their imagery scores with half being high imagers and the other half low imagers. Where all variables were incorporated, the design implied a mixed analysis of variance with three between-subjects variables and one within-subjects variable.

Subjects

The subjects were 160 undergraduate students enrolled in the introductory educational psychology course at The Pennsylvania State University. They received credit toward their course grade for participating in the experiment.

Selection of Imagery Groups

Prior to the conduct of the experiment proper, 345 subjects were administered the Space Thinking (Flags) test (Thurstone & Jeffrey, 1959), the Space Relations test of the Differential Aptitude Test Battery (Bennett, Seashore & Wesman, 1963), and the Gottschaldt Figures Test as described by Thurstone (1944). A factor score for each subject was obtained following the procedure described by Glass and Maguire (1966) in which the raw test scores are weighted by their respective factor loadings. These loadings were obtained from a factor analysis conducted previously and described in an earlier report by Di Vesta, Ingersoll, & Sunshine (1971, in press). The 80 subjects with the highest factor scores were characterized as the high imagers and those 80 with the lowest scores were characterized as low imagers. The subjects within each of these groups were randomly assigned to each of the four experimental treatments involved in the mnemonic aid variable crossed with the mediational mode variable. The only restriction in the random assignment of subjects was that there would be an equal number of subjects (n = 20) in each cell of the design.

Materials

Serial learning lists. The 40 words used to construct the lists for the serial learning tasks were selected from the concreteness (C), imagery (I), and meaningfulness (*m*) norms reported by Paivio, Yuille, & Madigan (1968). Of the words selected, 20 were abstract and low on rated imagery and 20 were concrete and high on rated imagery. The two groups of words were equivalent in *m*. The means of the different attributes for the concrete (C) and abstract (A) lists, respectively, were: $\overline{X} = 6.54$ and $\overline{X} = 2.90$ for imagery-ratings; $\overline{X} = 6.91$ and $\overline{X} = 2.11$ for concreteness-ratings; and $\overline{X} = 6.13$ and $\overline{X} = 5.50$ for *m*, each of these latter two means is SD = ± 0.3 from the mean *m* value of all words in Paivio, et al.'s norms. Two lists, of ten concrete nouns in each list were constructed by randomly selecting words from the initial list of 20 concrete nouns. A similar procedure was employed in constructing two lists of ten abstract nouns. The lists are presented in Table 1.

Table 1

Rated Imagery, Concreteness, and m-values for all Words

In the Concrete and Abstract Serial Learning Lists

Word	Imagery Latency	Concreteness Rating	Meaningfulness (m)
List 1 - Concrete			
Piano	6.70	6.85	6.40
Candy	6.63	6.56	6.39
Accordian	6.50	7.00	5.89
Steamer	6.53	6.94	6.32
Dress	6.53	6.93	5.68
Elbow	6.30	6.94	5.16
Mule	6.60	6.96	6.12
Cigar	6.80	6.96	6.22
Frog	6.73	6.96	6.56
Macaroni	6.47	7.00	5.48
Mean	6.58	6.91	6.02
List 2 - Concrete			
Library	6.73	6.87	6.40
Skull	6.47	6.96	6.64
Tweezers	6.57	6.93	5.80
Engine	6.33	6.76	6.08
Corpse	6.50	6-89	6.52
Building	6.40	6.94	5.48
Headlight	6.43	6.90	6.32
Pipe	6.43	6.90	6.20
Leopard	6.77	7.00	6.83
Nail	6.50	6.96	6.08
Mean	6.51	6.91	6.24
Average of Concrete Lists	$\overline{X}_I = 6.54$	$\overline{X}_C = 6.91$	$\overline{X}_m = 6.13$

Table 1a

Rated Imagery, Concreteness, and m-values for all Words

In the Concrete and Abstract Serial Learning Lists

Word	Imagery Latency	Concreteness Rating	Meaningfulness (m)
List 3 -- Abstract			
Adversity	2.80	2.03	5.06
Belief	2.73	1.55	5.24
Ego	2.90	1.93	5.72
Irony	2.83	2.10	5.25
Rating	2.60	2.66	5.12
Hypothesis	2.40	2.25	5.36
Emancipation	3.20	2.49	5.20
Deceit	3.30	1.66	4.92
Exclusion	2.80	2.41	5.32
Ability	2.67	2.03	5.60
Mean	2.82	2.11	5.28
List 4 - Abstract			
Word			
Crisis	3.43	2.81	5.44
Mercy	3.40	1.59	5.20
Satire	3.37	2.33	5.67
Magnitude	2.50	3.03	5.68
Knowledge	2.97	1.56	6.36
Perception	3.17	2.33	5.80
Democracy	2.47	1.79	5.72
Intellect	2.93	1.83	5.56
Welfare	3.17	2.35	6.16
Chance	2.50	1.51	5.61
Mean	2.99	2.11	5.72
Average of Abstract Lists	$\bar{X}_I = 2.90$	$\bar{X}_C = 2.11$	$\bar{X}_m = 5.50$

Jingle words. The concrete and abstract jingle words were selected from a pool of 25 words rhyming with the numbers one to ten. The initial pool of words was obtained by searching two dictionaries (Stillman, 1965; Wood, 1936) of rhyming words. The nouns were rated by 15 subjects for the ease with which they evoked sensory images on a seven-point scale (Paivio, et al., 1968). Concreteness was also rated on a seven-point scale bounded by the terms abstract and concrete. Concreteness was defined in terms of the directness of sensory reference as used by Spreen & Schulz (1966). The production of data regarding meaningfulness (*m*) was accomplished by requiring 15 subjects to associate as many words as possible, within 30-secs., to each noun. The procedure described by Noble (1952) was used to obtain the *m* values for these words.

The mean C and I ratings of the nouns selected for the jingle with concrete pegwords were $\overline{X} = 6.06$ and $\overline{X} = 6.72$, respectively; while the mean C and I ratings of the nouns selected for the jingle with the abstract pegwords were $\overline{X} = 1.86$ and $\overline{X} = 2.52$, respectively. The mean *m* of the concrete nouns was 6.05 and of the abstract nouns it was 5.95. Thus, the jingle words differed on the basis of imagery and concreteness but were essentailly the same on the basis of meaningfulness.

The words selected for the jingle comprised of concrete nouns were: one-bun; two-shoe; three-tree; four-door; five-hive; six-sticks; seven-leaven; eight-gate; nine-wine; and ten-hen. The words selected for the jingle comprised of abstract nouns were: one-fun; two-review; three-spree; four-chore; five-tithe; six-rhetoric; seven-heaven; eight-fate; nine-divine; ten-amen.

Procedure

The subjects were administered the experimental tasks in groups of two to four. At the outset, they were instructed in the reporting of verbal and imaginal mediation. Each person was given a stopwatch and instructed on its use in order to time himself at each task. After practice in using the stopwatch for two to three minutes, they memorized either the concrete or abstract jingle until the criterion of two successively correct recitations without error was reached. The experimenter then read aloud one of two sets of instructions depending upon the treatment being administered, i.e., the subjects were instructed to employ an imagery-set or a verbal set in performing the tasks. In the use of the imagery-set the subjects were instructed to connect, with a mental picture or image, each noun in the list to be learned with the jingle noun in the corresponding serial position and to record his connection by drawing a picture, however crude it might be. The verbal-set required subjects to relate, by forming a sentence or phrase, each noun in the list to the jingle noun in the corresponding serial position and to record his connection in verbal form. Subjects were reminded, before each list to be memorized had been read, to reset their stopwatches.

Following the preliminary instructions, the subjects were given two concrete noun-pairs and one abstract noun-pair in order to practice the use of the jingle and mediational set. The experimenter then read aloud the first noun in the list, and instructed the subjects to "start" at which time the watches were started. The subject stopped the watch as soon as he formed the complete mental image or verbalization. After the connection was recorded the subject recorded the time to the nearest second needed to form the link-up. The subjects were tested for serial recall after each of the four lists was presented.

The presentation of the four lists were counterbalanced among subjects by the use of a simple Latin-square to minimize the possible effects on recall of the order in which the lists were presented. After the recall test for the last list the experimenter asked the subjects to recall all of the words from the four lists that were in the first position, all of the words from the four lists that were in the second position and so on to the words in the tenth position, to determine how many of the 40 words the subject could retrieve.

Results

Several measures of performance were obtained including number of concrete- and abstract-word omissions after each list was memorized, intrusions from one list to another in recall sessions immediately after each list, omissions in the final recall task, intrusions from one list to another in the final recall task, and latencies in arriving at an association during the memorization task itself. Overall inter-correlations between the imagery-ability measure and each of these dependent variables indicated relatively high interrelationships among the measures. Accordingly, it was decided that the most efficient procedure was to perform analyses of the latencies in arriving at an association between the mnemonic aid (jingle) and the words in the list to be memorized; total errors, separately for concrete and abstract lists, made during the recall tasks after each list; and total errors of both types made on the final recall task. Mixed analyses of variances were made of each measure. In each analysis the between-subjects variables were level of imagery-ability (high or low imagers), kind of mediator (pictorial or verbal), kind of mnemonic aid (concrete or abstract pegwords). The within-subjects variable was kind of list (concrete words or abstract words) to be learned. The analyses based on these measures are summarized in Table 2. Since a triple-interaction involving imagery-ability, mediational set, and kind of list was obtained in the analysis of errors made on the final recall task, separate factorial analyses of variance, based only on the between-subjects variables, were routinely made for error scores on the concrete lista and for error-scores on the abstract lists of words. The results of these subanalyses are reported below only where it seemed necessary to do so in order to clarify the locus of a given effect.

Latency Measures

The initial analysis involved the measure of time, in seconds, to arrive at an association between the pegword in the mnemonic aid and the corresponding word in the list to be memorized. This analysis yielded $\underline{F}$ (1,152) = 18.01, $\underline{p}$ < .01, for the main effect due to the kind of mediational set; $\underline{F}$ (1,152) = 10.11, $\underline{p}$ < .01 for the main effect due to the kind of mnemonic aid; and $\underline{F}$ (1,152) = 14.99, $\underline{p}$ < .01 for the main effect due to kind of list memorized.

The time taken to arrive at an association by subjects who were to use a picture (imaginal set) in linking the pegword with a list word was longer on the average ($\overline{X}$ = 138.14 secs.) than that taken by subjects who linked the two sets of words via a sentence or verbal set ($\overline{X}$ = 99.69 secs.). It took less time to link each of the concrete words ($\overline{X}$ = 104.52 secs.) with its corresponding pegword than it did to link each of the abstract words ($\overline{X}$ = 133.32 sec.).

Table 2

Summary of Analyses of Variance of Recall Errors

Source	df	Latency[a] MS	Latency[a] F	Errors Immediate Recall MS	Errors Immediate Recall F	Errors Final Recall MS	Errors Final Recall F
Between-subjects							
Imagery (A)	1	18942	2.89	147.15	8.97*	277.51	11.18*
Mediators (B)	1	118272	18.01*	.70	.04	4.05	0.16
Mnemonic (C)	1	66355	10.11*	302.25	18.42*	515.11	20.76*
A x B	1	7527	1.15	169.65	10.34*	357.01	14.39*
A x C	1	22916	3.49**	11.63	0.71	2.45	0.10
B x C	1	588	0.09	104.65	6.38**	143.11	5.77**
A x B x C	1	5461	0.83	48.83	2.98	115.20	4.64**
Error (b)	152	6566		16.41		24.82	
Within-subjects							
Lists (J)	1	25920	14.99*	580.50	64.76*	1748.45	256.57*
A x J	1	2464	1.43	0.08	0.01	2.11	0.31
B x J	1	2565	1.48	24.75	2.76	20.00	2.94
C x J	1	610	0.35	0.53	.06	78.01	11.45*
A x B x J	1	485	0.28	2.28	.25	37.81	5.55**
A x C x J	1	262	0.15	0.38	.04	4.50	0.07
B x C x J	1	2952	1.71	1.38	.15	3.12	0.05
A x B x C x J	1	768	0.44	1.40	1.57	0.00	0.00
Error (w)	152	1729		8.96		6.81	

[a] Decimal places have been dropped for mean squares of analysis of latency scores.

* $p < .01$

** $p < .06$

The effect due to the interaction between imagery ability and kind of mnemonic aid used yielded $\underline{F}$ (1,152) = 3.49, $\underline{p}$ < .06. The means related to the interaction between imagery-ability and kind of mnemonic aid employed indicated the difference in time taken by high imagers ($\overline{X}$ = 105.29 secs.) and low imagers ($\overline{X}$ = 103.75 secs.) when using the concrete mnemonic was not significant. However, when the high imagers used the abstract mnemonic they took an average of 117.16 secs. for linking each words to its conceptual peg while the low imagers required an average of 149.47 secs. This finding is consistent with predictions form Paivio's two-stage association model.

The main effect due to imagery and the remaining interactions were found not to be significant ($\underline{p}$ > .10).

<u>Errors on the Immediate Recall Task</u>

The total number of errors made on recall after each list were analyzed via a mixed analysis of variance. The score for the number of errors was based on the summed omissions and intrusions. Thus, the score represents the exact opposite of the number of correct responses. This analysis indicated that the difference between imagery - ability groups was significant, $\underline{F}$ (1,152) = 8.97, $\underline{p}$ < .01; that there was a significant main effect due to mnemonic aids, $\underline{F}$ (1,152) = 18.42, $\underline{p}$ < .01; and a significant main effect due to kind of list memorized, $\underline{F}$ (1,152) = 64.76, $\underline{p}$ < .01. In addition, the analysis revealed significant interactions between imagery-ability and kind of mediator, $\underline{F}$ (1,152) = 10.34, $\underline{p}$ < .01; between mediators and mnemonic aids, $\underline{F}$ (1,152) = 6.38, $\underline{p}$ < .05.

The differences in mean number of errors among groups indicated in the aforegoing analyses, are summarized in Table 3. In brief, these data indicate that high imagers made significantly fewer errors ($\bar{X} = 3.36$) than low imagers ($\bar{X} = 6.17$) when imaginal sets were used. However, the difference in means for the two groups ($\bar{X} = 4.73$ and $\bar{X} = 4.62$ for high and low imagers, respectively) was not significant when verbal mediators were used. Additionally supportive of the notion that imagery facilitates learning and recall is the finding that fewer errors were made with concrete mnemonic aids ($\bar{X} = 3.75$) than with abstract mnemonic aids ($\bar{X} = 5.69$) and that this difference is considerably greater when imaginal mediators (difference = 3.09) were used than when verbal mediators (difference = 0.79) were used. There is a hint in this analysis of the interaction between imagery and treatments which appears strongly in the analysis of the final recall data presented immediately below. Thus, in the present analysis high imagers made fewer errors ($\bar{X} = 2.40$) than low imagers ($\bar{X} = 4.05$) when imaginal mediators were used with concrete mnemonic devices, $t = 1.28$, $p < .10$. Low imagers were especially hindered ($\bar{X} = 8.30$) when using imaginal mediators and abstract mnemonic devices conjunctively compared to high imagers ($\bar{X} = 4.33$), $t = 3.09$, $p < .01$. When using the verbal set with concrete mnemonic aids high imagers tend to make slightly fewer errors ($\bar{X} = 4.13$) than low imagers ($\bar{X} = 4.42$); and when a verbal set is used with abstract mnemonic-aid high imagers make more errors ($\bar{X} = 5.33$) than do low imagers ($\bar{X} = 4.83$). However, the interaction represented by these means is not significant ($p > .10$).

Table 3

Summary of Mean Number of Errors Made on the Recall Task, Immediately After Each List Presentation, by High and Low Imagers Employing Concrete and Abstract Mnemonics Under Conditions of Imaginal and Verbal Mediation

Imagery-Ability	Kind of Mnemonic: Concrete Mnemonic	Abstract Mnemonic	Overall
Imaginal-set			
High Imagers	2.40	4.33	3.36
Low Imagers	4.05	8.30	6.17
Total	3.22	6.31	4.76
Verbal-set			
High Imagers	4.13	5.33	4.73
Low Imagers	4.42	4.83	4.62
Total	4.28	5.07	4.67
Overall	3.75	5.69	4.72

Errors on the Final Recall Task

The analysis of errors in the final recall task was based on the score involving all incorrect responses and omissions made in recalling words in given positions from all lists after the four lists had been memorized. In this analysis the effect due to imagery-ability yielded F (1,152) = 11.18, $p < .01$; that due to kind of mnemonic aid yielded F (1,152) = 20.76, $p < .01$; and the effect due to lists yielded F (1,152) = 256.67, $p < .01$. Of the first order interactions three were found to be significant. These were F (1,152) = 14.39, $p < .01$, for the interaction between imagery-ability and kind of mediational set; F (1,152) = 5.77, $p < .05$ for the interaction between mediational set and mnemonic aids; and F (1,152) = 11.45, $p < .01$ for the interaction between kind of mnemonic aids and kind of lists. These main effects and interactions must be further qualified by the significant second-order interaction between imagery-ability, mediational set and mnemonic aids which yielded F (1,152) = 4.64, $p < .05$, and between imagery-ability, mediational set and kind of list which yielded F (1,152) = 5.55, $p < .05$.

The mean number of errors for these effects are summarized in Tables 4 and 5. The reader will note that high imagers made fewer errors ($\bar{X}$ = 5.60) than low imagers ($\bar{X}$ = 9.60) when using imaginal mediators. On the other hand, there is no difference between the two groups when verbal mediators are used, high imagers made an average of 7.96 errors while low imagers averaged 7.71 errors. The performance of high imagers was particularly facilitated when they used imaginal mediators with concrete mnemonic aids ($\bar{X}$ = 4.37) and concrete words ($\bar{X}$ = 3.80). The low imagers were hindered to a considerable degree when they employed imaginal mediators with the abstract mnemonic aid

Table 4

Summary of Mean Number of Errors Made During Final Recall

By High and Low Imagers Employing Imaginal and Verbal Mediators

In Learning Concrete and Abstract Lists

Imagery Ability	Kind of List		Row Means
	Concrete	Abstract	
Imaginal-mediator			
High Imagers	3.80	7.45	5.60
Low Imagers	7.25	11.95	9.60
Sub-means	5.53	9.70	7.61
Verbal-mediator			
High Imagers	4.95	10.98	7.96
Low Imagers	5.58	9.88	7.71
Sub-means	5.25	10.43	7.83
Column means	5.39	10.06	7.72

Table 5

Summary of Mean Number of Errors Made During Final Recall

By High and Low Imagers Employing Concrete and Abstract Mnemonics

With Imaginal and Verbal Mediators

Imagery Ability	Kind of Mnemonic: Concrete Mnemonic	Abstract Mnemonic	Row Means
<u>Imaginal</u> <u>Mediator</u>			
High Imagers	4.37	6.87	5.63
Low Imagers	6.97	12.22	9.60
Sub-means	5.67	9.55	7.61
<u>Verbal</u> <u>Mediator</u>			
High Imagers	6.85	9.08	7.96
Low Imagers	7.63	7.80	7.71
Sub-means	7.24	8.43	7.83
Column means	6.46	8.99	7.72

($\bar{X}$ = 12.22) and with the abstract word-list ($\bar{X}$ = 11.95). Multiple comparisons made of the data presented in Tables 4 and 5 via the t-test indicate only the differences between high and low imagers employing the imaginal mediator with abstract lists (t = 3.31) or with abstract mnemonics (t = 3.38) were significant (p < .01). The interaction between imagery ability and kind of list in the verbal mediation condition yielded F = 1.87, p < .065.

In general, high imagers always made fewer errors than did low imagers when imaginal mediators were used. High imagers retained this advantage, although to a considerably lesser degree, when verbal mediators were used with the concrete mnemonic and concrete word list. However, the low imagers made fewer errors than high imagers when learning the abstract word lists or when using the abstract mnemonic when verbal mediators were used. Thus, while the analysis of results on delayed recall are in general agreement with those obtained for immediate recall, the differences obtained are larger. Accordingly, it appears that one condition for identifying aptitude by treatment interactions is the examination of its influence of imagery, under conditions comparable to those in the present experiment, over longer delay periods, perhaps a delay of a week or two.

Discussion

The results of this study clearly imply that rated imagery (i.e., concreteness) as a stimulus attribute, imaginal strategy as a mediational process, and imagery-ability as an individual difference variable are similarly related to performance and recall. Thus, pairs of concrete nouns are acquired more easily than pairs of abstract nouns, an imaginal mediational strategy yields fewer errors than a verbal mediational strategy, and learners with high imagery-ability produce fewer errors than those with low imagery-ability.

The critical comparisons for this study, however, involved the interactions among these variables, particularly the performance of high and low imagers when learning word-pairs comprised of abstract stimuli or when using abstract mnemonics under sets to employ imaginal rather than verbal processes. With respect to these relationships it was found that for the immediate recall task, the high imagers performed significantly better than low imagers when the abstract mnemonic device was combined with the imaginal set. In no other condition was the high imagery group superior to the low imagery group on this task. The same result was obtained in the final (delayed) recall task. The concreteness of words within lists as a variable interacting with imagery ability was not supported when imaginal mediators were used; i.e., the low imagers made significantly more errors than high imagers on both kinds of lists when imaginal mediators were used. However, the interaction between imagery ability and kind of list on the final recall task indicated a tendency for high imagers to perform better on the concrete task and low imagers to perform better on the abstract task when verbal mediators were used. Furthermore, in all comparisons with either abstract word lists or mnemonic devices with abstract pegwords, low imagers performed significantly ($p < .01$) better with verbal sets than with imaginal sets.

It would appear that the two-stage model requires, and has received in this study, both kinds of support when comparing the performance of high and low imagers. Thus, high imagers when learning abstract stimuli impose an imagery strategy on the material being learned. When required (i.e., forced) to use this strategy they can perform effectively. However, because they typically employ an imaginal strategy in processing information, if required to use another strategy, such as a verbal one, they perform much more poorly than the low imagers who, presumably, employ other strategies to better advantage than the imagery one. Conversely, the low imager who does not employ imagery to good advantage has difficulty in using the imaginal set, in fact his performance is seriously hampered when compared to his performance under the verbal set. Thus, the hypotheses regarding the two-stage model and the role of imagery ability in associative learning are provided considerable support in the present study.

A major difficulty in conducting studies on trait by treatment interactions appears to be in determining how the trait is to be measured. Thus, in the present study, high imagers perform about as expected. However, low imagers are not really to be considered verbalizers. It is not clear what their dominant strategy is except that they are people deficient in some strategy (i.e., in imagery-ability) without knowing their strengths. Yet some means of identifying a group with a strategy that is "opposite" to imagery (verbalizers?) in the same way that imaginal processes are "opposed" to verbal processes is clearly required. Upon analysis, identification of a strategy such as that of verbalizing may be found to be a difficult task. Are verbalizers learners who are verbally fluent? with excellent vocabularies? with flawless language habits? with good reading comprehension? a composite of all of these? A general verbal trait or ability can be easily identified; a trait closely related to the acquisition of associates may be a much more difficult task.

Nevertheless, the overall results of this study clearly indicate that whatever has been measured by the battery of "imagery" tests is definitely related to the processing of information by the learner. Thus, the major distinction between high and low imagers is at least in their ability, or preference, to use imagery over some other learning-recall strategy. When conditions favor the use of this strategy the high imager is at an advantage. This general conclusion is supportive of the two-stage model of associative learning and is comparable to a result reported by Yuille & Paivio (1967) who found that mediation latency was unrelated to stimulus concreteness and mediation set when concrete stimuli were employed but that imagery set was significantly inferior to verbal set when abstract stimuli were used. In the present study the parallel comparison is of the performance of low imagers employing imaginal and verbal sets with abstract mnemonics or lists. In each of these comparisons, the performance of the low imager is hindered when he is forced to employ imaginal mediators, but is unimpaired when he is forced to use verbal mediators.

In summary, the reasoning and results of the present study indicate that the relationships between traits and treatments may be in the ability of the individual to deal with the task in general, his receptivity to, or preference for, certain kinds of stimuli over others, or the strategies by which he attacks a task or processes the information. Stewart (1965) and Hollenberg (1970) both assumed that imagery-ability affected the receptivity to stimuli and accordingly their investigations compared the acquisition and recall of learners when presented pictorial and verbal stimuli. Without minimizing the importance of presentation mode as a variable in learning, the present study suggests that the manner in which the learner processes the material, in terms of the task requirement (also see Ernest & Paivio, 1969), is as important as the effects of manner of presentation. For an understanding of the dynamics of the learning process including the role of individual differences on performance, it may be more important.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential aptitude tests. (Grades 8-13 and adults.) New York: Psychological Corporation, 1963.

Di Vesta, F. J., Ingersoll, G., & Sunshine, P. A factor analysis of imagery tests. Journal of Verbal Learning and Verbal Behavior, 1971, in press.

Ernest, C. H., & Paivio, A. Imagery ability in paired-associate and incidental learning. Psychonomic Science, 1969, 15, 181-182.

Ernest, C. H., & Paivio, A. Imagery and verbal associative latencies as a function of imagery ability. Canadian Journal of Psychology/ Review of Canadian Psychology, 1971, 25, 83-90.

Glass, G. V., & Maguire, T. O. Abuses of factor scores. American Educational Research Journal, 1966, 3, 297-304.

Hollenberg, C. K. Functions of visual imagery in the learning and concept formation of children. Child Development, 1970, 41, 1003-1015.

Noble, C. E. An analysis of meaning. Psychological Review, 1952, 59, 421-430.

Paivio, A. Learning of adjective-noun paired associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370-377. (a)

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263. (b)

Paivio, A., & Foth, D. Imaginal and verbal mediators and noun concreteness in paired-associate learning: the elusive interaction. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 384-390.

Paivio, A., & Yuille, J. C. Changes in associative strategies and paired-associate learning over trials as a function of word imagery and type of learning set. Journal of Experimental Psychology, 1969, 79, 458-463.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, 76, No. 1, Part 2.

Spreen, O., & Schulz, R. W. Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 459-468.

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, University of Toronto, 1965.

Stillman, F. The Poet's Manual and Rhyming Dictionary. Thomas Y. Crowell Company, New York, 1965.

Thurstone, L. L. A factorial study of perception. Chicago: University of Chicago Press, 1944.

Thurstone, L. L., & Jeffrey, T. G. Space thinking (Flags). Chicago: Education-Industry Service, 1959.

Wood, C. The Complete Rhyming Dictionary and Poets Craft Book. Halcyon House: New York, 1936.

Yuille, J. C., & Paivio, A. Latency of imaginal and verbal mediators as a function of stimulus and response concreteness-imagery. Journal of Experimental Psychology, 1967, 75, 540-544.

Summary

The Effect of Context Modality on Acquisition and Transfer By Imagers and Non-Imagers*

Study Director: Steven M. Ross

Advisor: Francis J. Di Vesta

Technical Problem

This study investigated the conditions under which individuals who differ in imagery abilities acquire and transfer concepts that are incidentally expressed by pictorial or verbal contexts. It has been shown in an earlier investigation (Di Vesta & Ross, 1970) that the relatedness or meaningfulness of a verbal context has demonstrable effects upon the learning and transfer of paired-associates. Specifically, it was determined that a related context, i.e., one which is similar in meaning or categorically relatable to the stimulus side of the pair, elicits conceptualizing tendencies in the learner that interfere with specific item learning but facilitate conceptual transfer.

The present study extended the earlier one by manipulating the modality of the context, and also by incorporating imagery ability as an individual difference variable. It was hypothesized that since imagers would be more receptive to pictorial than to verbal contextual cues,

* This is an abstract of a master's thesis in preparation at the time the present report was being prepared.

they would acquire the pictorially expressed concept more readily than non-imagers. Non-imagers, on the other hand, were expected to be less influenced by context modality and therefore able to learn concepts expressed either pictorially or verbally with equal facility.

General Methodology

The experiment consisted of two phases: learning and transfer. During the learning phase, subjects learned a list of paired-associates to a criterion of one errorless trial. In one condition, two context words that were meaningfully related to each other and to the specific stimulus elements, were placed between the stimulus and response positions of all word-pairs. In a second condition the context was identical in meaning and similarly positioned as in the first condition but was presented pictorially. In a third condition no context of any kind was present. All stimuli were mounted on slides and presented by means of a carousel projector.

During the transfer phase, subjects learned a new list of paired-associates. Each word-pair in the list was comprised of the original response element from the learning phase list and a new stimulus. Context words were eliminated in the transfer phase. In one set of conditions the new stimulus word was categorically related to the original stimulus and to the context presented during the learning phase. In the other set of conditions the new stimulus was related only to the original stimulus, but not to the context. Thus, depending upon condition, subjects were required to transfer to a concept that was either identical or alternate to that of the learning context.

Technical Results

The analysis for the learning phase of the experiment indicated that the aptitude-by-treatment interaction involving imagery ability and context modality was significant. The direction of the data revealed that there were no differences in recall between imagers and non-imagers in the verbal context condition but that imagers were highly superior in both the picture context and no-context conditions. There were no significant differences overall between the context variations.

The analysis of data for the transfer phase indicated that performance on the same-concept transfer task was significantly better than performance on the alternate-concept transfer task. More important, however, was the finding that this difference was not pronounced when the pictorial context, as opposed to verbal context, was employed in the learning phase. Imagery ability did not significantly interact with these factors though the positive effects of the picture context tended to be slightly greater for imagers than they were for non-imagers.

Educational Implications

The present investigation involved the assumption that some individuals, more than others, habitually employ concrete images in the encoding of informational inputs. Accordingly, it appears likely that these individuals would demonstrate a greater capacity for learning from concrete kinds of materials than they would from materials that were relatively symbolic or abstract.

Though the present findings cannot be interpreted as unconditionally supportive of the above notion, they clearly imply that modality of presentation and relatedness of incidental cues provide constraints on the storage and transfer of incoming information. It can also be inferred from this study that, at least during acquisition, high imagers are more receptive to and more able to effectively process information that is embedded within a pictorial context than are low imagers. This suggests that students who demonstrate a preference for coding via imagery might best benefit from a context of concrete examples and graphic displays. If this interpretation proves to be viable then it implies a definite responsibility for the teacher to adapt the mode of the instruction to the individual preference and ability of the learner provided economy of acquisition is an important objective.

Implications for Further Research

In the present study, the effects of imagery as a stimulus attribute and as an individual difference variable were examined as determinants of learners' abilities to acquire and transfer information. The differential performances of imagers and non-imagers under the influence of pictorial and verbal contexts implies the need to investigate further the adaptation of instructional techniques to this kind of individual difference. Of considerable importance would be the assessment of the generalizability of the above findings and the degree to which they are translatable to applied classroom practices. Further research might also suggest methods of instruction that provide remedial assistance for individuals who demonstrate an excessive reliance upon visual strategies of processing information.

References

Di Vesta, F. J., & Ross, S. M. Contextual cues and cognitive structures in the storage and retrieval of information. Semi-annual report submitted to the Advance Research Projects Agency, 1970. Pp. 68-97.

Section II

Organization of Materials and

the Learner's Instrumental Activities

PRECEDING PAGE BLANK

Summary

Listening and Note-taking II: Thematic Content, Note-Taking, and Length of Listening Review Intervals as Variables Influencing Immediate and Delayed Recall*

Francis J. Di Vesta and G. Susan Gray

Technical Problem

This study was based on the consideration that the recall of material from a lecture may be influenced by three variables: note-taking, length of the listening-study interval, and thematic organization (i.e., relatedness of the content of the lecture). Underlying these operations were assumed to be the facilitative functions of external storage and transformation of material served by note-taking; consolidation of material by frequent interruptions during which the learning could "mature"; and, linking of the new material to previously learned material when the content was related.

General Methodology

The subjects were presented a 30-minute passage via a recording. Passages were broken into 5-minute segments each of which contained the same number of words and ideas. The contents of these segments were either all related to a common theme; related to a common theme but on different topics; or on completely different topics. These variations

---

* This is a preliminary report. Other analyses being made of the data were not completed at the time the technical report was being prepared.

PRECEDING PAGE BLANK

were completely crossed with three variations of listen-study intervals (i.e., 5-minute - 2 minute; 15-minute - 6 minute, and 30-minute - 12 minute combinations) and two variations of note-taking (i.e., note-taking permitted or note-taking not permitted).

Technical Results

The findings indicated that note-taking and thematic relatedness were related to the ability of subjects to recall material to which they had listened. More ideas were recalled when notes were taken than when note-taking was not permitted and more ideas were recalled when the material was on different topics than when the material was all related to the same topic. These effects were especially noted on a delayed recall test administered one week following the listening period. One interpretation of the finding that more is remembered from the discontinuous material is that less opportunity for interference exists in such material than it does in continuously related content. No effects were found due to variations in the listening-study intervals or the interaction between the treatments and the memory-span (i.e., the individual difference variable.)

Educational Implications

As Crawford (1925) had indicated in his study, it was also found in the present study that the immediate value of notes appears to be of less value than the delayed value. Notes appear to increase the saliency of given ideas and make them more subject to recall, though not necessarily guaranteeing that they will be recalled. These studies point clearly in the direction that note-taking acts more as a transformational aid for the student than as a mere external storage mechanism.

The findings regarding better recall of discontinuous material suggests that interruptions in ideas presented during a lecture might be beneficial. Stated somewhat differently, a lecturer might segment a lecture period by providing opportunity for discriminability among ideas presented, even though such ideas are distinctly related.

Implications for Further Research

The finding regarding the effect of thematic relatedness requires replication. Even though this finding parallels others found by Anderson & Carter (1971) it appears desirable that still other materials should be employed. One difficulty in conducting research with connected discourse is in equating all passages for such characteristics as familiarity and meaningfulness. An alternative, of course, is to randomize passages. Variations in content can also be accomplished by changing the degree of discriminability from one treatment to another as implied in the aforegoing section. Finally, it is suggested that future investigations of note-taking should be based on an analysis of the processes or strategies underlying this device much as has been done in Project Ikon (see Table of Contents for page numbers).

References

Anderson, R. C., & Carter, J. Retroactive inhibition of meaningfully-learned sentences. Mimeographed technical report, University of Illinois, 1971.

Crawford, C. C. Some experimental studies of the results of college note-taking. Journal of Educational Research, 1925, 12, 379-386.

Listening and Note-taking II: The Effects of Variations in Thematic Continuity, Note-taking, and Length of Listening-Review Intervals on Immediate and Delayed Recall*

Francis J. Di Vesta and G. Susan Gray

In a previous study by Di Vesta and Gray (1971) it was found that the number of ideas recalled by subjects after listening to a short passage was favorably influenced by note-taking, by rehearsal immediately after listening to a passage, and by testing on the passage. In comparison a study by Eisner & Rohde (1959) led to the conclusion that taking notes during a 3-min. lecture is not superior to delayed note-taking (i.e., note-taking immediately after the lecture). However, the differences between results of the two studies can easily be attributed to the consequence of procedural differences. Thus, for example, Eisner & Rohde had some subjects take notes and then study them immediately after listening while other subjects were to concentrate on the lecture and then jot down as many notes as they could, immediately after listening. In the Di Vesta & Gray study the subjects in one treatment used the rehearsal period to study their notes, or to contemplate what they had heard, and in another treatment they were prevented from rehearsing the material.

An analysis of the aforementioned studies together with more subjective considerations of the note-taking process suggests two other variables that may affect recall of material when notes are being taken during a lecture: the length of a consolidation period and the degree of thematic organization of material presented in the lecture.

A consolidation period has been suggested by M. A. Howe (1970) and Hebb (1966) as a period during which the material can be rehearsed or coded for more efficient storage in memory. More specifically, Hebb (1966, pp. 122-123) suggests that "... newly acquired learning must be undisturbed for some time if it is to last ... and must be allowed to mature ... between 15 minutes and an hour or thereabouts." He speculates, too, that reinforcement may strengthen learning because it provides for a period during which consolidation can occur. On the basis of such assumptions it would appear that there are several opportunities during presentation of a lecture for providing a consolidation period. Thus, the lecturer himself, may pause to allow the material to "sink-in" or note-taking as an activity may function to direct the student's attention to certain parts of the material, perhaps at the expense of attention to other parts, but in the process allowing the important points to "mature". The most salient opportunity for consolidation may be a period immediately following a lecture during which time the student can review and contemplate all that had been said.

A somewhat neglected variable that appears to account for differences in procedures among studies on note-taking is what the present investigators have called the "thematic relatedness" of the lecture contents. Eisner & Rohde (1959) gave their subjects a 30-min. lecture on the short story and another on romanticism, each of which could be considered as self-contained passages on a single theme. Crawford (1925a; 1925b) employed a series of lectures in a college course; again these were probably more thematically related than not, although his descriptions do not permit a precise evaluation. Berliner (1970) used a 30-min. passage comprised of thematically related material in the sense that it described the history of China but could also be considered as segmented because it described several independent aspects (e.g., dynasties, mongols, kingdoms, and rebellions) of China's development. Di Vesta & Gray (1971) used three 5-min. passages each completely independent of the others. Obviously, then, thematic relatedness of the content involved in the passages to which the subjects listened differed among studies. The implication of such differences is that thematic relatedness may have an effect on the strategy the student uses to store the material during learning and his ability to recall it on a later occasion. It may also have an effect on recall either through proactive or retroactive inhibition (Wickens, 1970) or through the possibilities for organization of the passage. Ausubel (1968) has been a leading advocate of improving didactic methods of instruction through increasing meaningfulness for the learner. Such considerations imply that the period prior to consolidation should be filled with material characterized by a minimal degree of meaningfulness to permit coding. While the criterion by which meaningfulness is defined must necessarily remain vague until its characteristics can be more fully specified, the present investigators assumed that thematic relationship may be one such characteristic.

Another consideration, of importance in an analysis of note-taking, is the number of ideas presented sequentially prior to a consolidation period. It can be argued that, because of the limitations of short-term memory, a lecture that contains too many ideas in a single sequence (which would be comparable to massed practice) would be difficult to retain since there would be insufficient time for encoding (organizing) the material. A presentation (lecture) that was too long would also create difficulty for the learner because the material could not be rehearsed; it would not remain in short-term memory long enough to be encoded. Either situation would result in inefficient recall or retrieval, especially over the long-term.

In view of the aforegoing rationale, the present study was designed to extend previous studies on note-taking and listening by investigating the effects of thematic relatedness and of opportunity for consolidation on the learner's recall. Note-taking was incorporated into the design to provide continuity with earlier experiments and to determine whether note-taking interacts with the other variables.

In view of the fact that the present study emphasized memorial processes, it was assumed that individual differences in either coding ability or short-term memory would interact with the treatments considered. Due to limitations on time that could be demanded from subjects who participated in this study, both measures could not be administered. Accordingly, as an initial attempt at investigating aptitude by treatment interactions, it was decided to employ only a test of short-term memory on the reasoning that students with low short-term memory scores would have developed a strategy of taking notes as a way of compensating for their limited memory span. It was hypothesized that such persons would profit by taking notes, especially with an opportunity for review, and would be handicapped if prevented from doing so, especially without an opportunity for review, to a greater extent than would persons with higher short-term memory scores.

Experiment I

Method

Design. The subjects were randomly assigned to one of 18 conditions represented in a design consisting of three orthogonally crossed variables. The variables consisted of three levels of thematic relatedness, three variations of listen-study intervals, and two alternatives for note-taking.

Each thirty minute lecture to which the subjects listened was comprised of 5-min. segments. To achieve different levels of thematic relatedness, the degree to which the six 5-min. segments were associated to a common topic was varied. In the continuous related thematic (CRT) condition, the topic of the 30-min. passage was an historical narrative, segmented at 5-min. intervals. In the condition where the material was discontinuous but related by a common theme, each 5-min. passage covered a distinctly different topic, but the six topics were conceptually related to a superordinate idea (Condition DRT). The third treatment was called discontinuous-unrelated (DUR) since it was comprised of six passages whose topics were completely unrelated to one another.

Each of the three 30-min. thematic conditions, being separable into 5-min. segments, were manipulated to achieve 5-min., 15-min., and 30-min. variations in the length of the listening interval. In one variation of the listen-study interval, each 5-min. segment was followed by a 2-min. study interval, until all six segments had been presented. In the second variation, three of the 5-min. segments were presented consecutively, followed by a 6-min. study period, thereby resulting in two sequences consisting of a 15-min. listening period followed by a 6-min. study period. The third variation of the listen-study interval consisted of the consecutive presentation of all six 5-min. passages, producing a 30-min. listening period, followed by a 12-min. study period. Thus, all groups had equal amounts of listening and study time. In a sense, these treatments were comparable to variations in massed and distributed practice.

The third set of treatments consisted of variations in the opportunity to take notes. Half of the subjects were provided with a booklet in which to take notes while listening to the passages; the other half of the subjects were instructed that they could not take notes. Consequently, during the study intervals, those subjects who took notes, could study their notes at this time. Subjects who were prohibited from taking notes used the study interval for mental rehearsal and recapitulation of the material to which they had listened.

Materials. The material for the continuous-theme was edited from Rachel Carson's Silent Spring (1962). The topic traced the history of the fire ant in the United States. In the DRT treatment, each of the six passages dealt with a distinct topic, a specific insecticide. Thus, while each 5-min. segment was an entity by itself, thematic relatedness was achieved by the inclusion of the topics of the six passages within the superordinate conceptual category of insecticides. The textual material was also edited from Carson (1962). The six segments of connected discourse employed in the DUR treatment dealt with six completely unrelated topics: the Chou dynasty in China (edited from Berliner, 1970); prison classification systems (Loveland, 1962); the Minnesota Theater (The Washington Post, 1970); and the insecticide, DDT (from Carson, 1962).

Each passage was constructed to meet as closely as possible the specifications of 500 words and 25 ideas. A separate tape was prepared for each of the three passages. The three different listen-study intervals were presented on separate tracks of each tape which was stored in the control room of a remote-deck library system language laboratory. The subjects, seated at individual booths in the language laboratory, could reach the tape to which they were assigned by means of a telephone dial.

Specific instructions were typed on bond paper and set on the desk in front of each subjects. If the subject was allowed to take notes, he was supplied with a booklet of note paper, one page for each of the six passages.

Subjects. The 90 subjects were volunteers for the experiment from an introductory educational psychology course at The Pennsylvania State University. They received credit toward their final grade in the course for their participation. None of the subjects had participated in an experiment requiring performance with connected discourse materials though most had participated in another experiment.

Procedure. A maximum of nine and a minimum of four subjects participated in the experiment during any one experimental session. The number of subjects varied because some subjects failed to appear at their assigned time. The subjects were randomly assigned to one of the 18 experimental conditions with the restriction that only one subject be placed in any one experimental cell during a single experimental session.

An additional restriction was imposed on the randomization procedure to assure that an equal number of subjects (n = 5) would be obtained for each cell.

When the subjects were seated in the experimental room the experimenter explained that the study was an attempt to investigate how students went about learning new materials. The operation of the laboratory equipment was also explained to the subjects at this time. After all questions were answered, the subjects put on their headphones and dialed the appropriate tape. When all were dialed in, the experimenter proceeded to the control room where the tapes were then started. After the last study period, the subjects were administered a free-recall test, followed by a 30 item true-false test on the material to which they had listened. For the free-recall test, the experimenter instructed the subjects to write down all they could remember about each passage in twenty minutes. They were told to use a separate page of the test booklet for each passage, but were also told that they could write on the passages in any order they wished--they did not have to follow the actual presentation order of the passages on the tapes. Following the 20 minute free-recall test period the subjects were administered the true-false tests.

One week after the initial experimental session, all subjects returned to take a 115 question true-false test, which included the 30 true-false test items comprising the initial test, and a memory span test patterned after Peterson and Peterson's short term memory task (1957).

Results

Separate 3 x 3 x 2 factorial analyses of variance were made of the number of ideas recalled in the free-recall task, the number of correct responses on the 30 item true-false test given immediately after the listening session, and the number of correct responses on the delayed true-false test.

The analysis of the number of ideas yielded $\underline{F}$ (2,72) = 24.42, $\underline{p} < .001$ for the effect due to the passage organization. The subjects remembered fewer ideas from the CRT material ($\overline{X} = 27.33$) than they did from the DRT material ($\overline{X} = 30.83$). The greatest number of ideas was recalled ($\overline{X} = 43.20$) by subjects who listened to the DUR material. Thus, the number of ideas recalled was influenced significantly by the thematic organization of the material. These results imply that recall was inversely related to the degree of thematic organization of the passages.

The effect due to the note-taking treatment yielded $\underline{F}$ (1,72) = 12.59, $\underline{p} < .001$. The subjects who were not allowed to take notes recalled an average of 30.33 ideas whereas the subjects who were allowed to take notes while listening recalled an average of 37.24 ideas.

None of the treatments significantly affected the subjects' performance on the immediate true-false test. However, the analysis of the scores for the delayed true-false test, given one week after the experimental session, yielded $\underline{F}$ (1,72) = 8.40, $\underline{p} < .005$ for the effect due to the note-taking treatment. Those subjects who were allowed to take notes achieved a mean score of 84.69 on the 115-item test, while subjects who took no notes achieved a mean score of 79.16. Although definite conclusions regarding functional relationships cannot be drawn from these data because of the differences in lengths, with consequent differences on reliability, of the immediate and delayed test, the data imply that performance on the delayed test was affected by the taking of notes.

Experiment II

The design used in Experiment I was extended in Experiment II by the addition of a fourth treatment to the thematic relatedness variable. The textual material for this treatment was identical to that incorporated in the continuous-related theme (condition CRT) in Experiment I, i.e., the history of the fire-ant (Carson, 1962). However, the material was not presented according to the logical and chronological order employed in Experiment I. Instead, the six segments were scrambled such that the 5-min. passage that was originally presented in the fourth position was presented first, followed by passages 2, 5, 1, 6, and 3, respectively. In addition, two forms of the true-false test were administered, one to half of the experimental groups and the other to the remaining groups, immediately after listening. The addition of the continuous-scrambled thematic (CST) treatment and the two forms of the true-false test resulted in an experimental design with four orthogonally crossed variables: four variations of thematic relatedness; three variations of length of listening interval; two note-taking treatments; and two forms of the test administered immediately after listening.

Method

Materials. The same materials were used as in Experiment I. A fourth tape was prepared for the continuous-scrambled theme, with three tracks for the three listen-study intervals, respectively.

Subjects. The subjects for this Experiment II were 240 students from another class of the introductory course in educational psychology. In all other respects their characteristics were essentially the same as those employed in Experiment I.

Procedure. The procedures for Experiment II were identical to those followed in Experiment I, except that the subjects in Experiment II were allowed as much time as they needed for the free-recall task. However, they were instructed that once they had begun the true-false test they could not return to the free-recall test.

As in Experiment I, all subjects returned one week after the experimental session to take the 115-item true-false test and the memory span test.

Results

In this experiment the free-recall material was scored for the number of ideas correctly recalled and the total number of words recalled. Each of these scores was separately analyzed by a 4 x 3 x 2 x 6 mixed analysis of variance in which the last variable was a within-subjects variable consisting of the six passages. The scores based on the true-false tests were analyzed by a 4 x 3 x 2 x 2 mixed analysis of variance in which the variables were the between-subjects factors of thematic organization, listening-review intervals, and note-taking and the within-subjects variable consisted of scores on the immediate and delayed tests. See Table 1 for a summary of this analysis.

Table 1

Summary of Analysis of Variance for Experiment 1

Source	*df*	MS	*F*
Thematic Organization (A)	2	2084.68	24.42*
Length of Listening-Review Interval (B)	2	50.41	.59
Note-taking (C)	1	1074.68	12.59*
A x B	4	24.38	.29
A x C	2	118.01	1.38
B x C	2	31.34	.37
A x B x C	4	34.08	.40
Error	72	85.36	

* $p < .001$

Effect of Treatments. The analysis of the number of ideas recalled yielded F (3,92) = 60.55, $p < .001$ for the effect due to thematic relatedness. The mean scores were: $\bar{X} = 54.54$ for the DUR group; $\bar{X} = 38.06$ for the DRT group; $\bar{X} = 33.31$ for the CST group; and $\bar{X} = 27.87$ for the CRT group. Thus, these findings are in substantial agreement with those obtained in Experiment 1. Furthermore, these results indicate that the additional time did allow for the elicitation of more correct ideas but that the advantage still favored the DUR group. Of interest in this regard is the finding that the CST material resulted in increased recall over the CRT material supporting the notion that there is an optimal amount of relatedness among paragraphs beyond which there is likely to be a debilitating effect on recall. A detailed summary of these results is presented in Table 2.

As in Experiment I, the effect due to note-taking yielded F (1,92) = 27.65, $p < .001$. The subjects who were allowed to take notes recorded a mean of 42.36 correct ideas while the group that did not take notes recorded 34.56 correct ideas on the average.

It is interesting to compare the results described in the aforegoing with the results obtained when the sheer number of words recalled is used as the dependent variable. The effect of thematic organization on this measure yielded F (3,216) = 21.77, $p < .001$. The differences related to this analysis are reflected in the means of 503.4 words used by the DUR group; 393.6 words by the CST group; 345.6 words by the CRT group; and 315.6 words by the DRT group. Thus, although the DUR group produced both the greatest number of ideas and words there was little relationship between the two variables in the results of the other groups.

Table 2

Number of Ideas Recalled From Each of the Segments

Of Each Passage According to Thematic Organization[a]

Continuous Related Theme		Continuous Scrambled		Discontinuous Related		Discont... ous Unrelated	
Segment	$\bar{X}$	Segment	$\bar{X}$	Segment	$\bar{X}$	Segment	$\bar{X}$
1	5.12	1	5.85	DDT	9.63	Chou Dynasty	8.02
2	4.73	2	4.43	Chlordane	4.88	Hypnosis	7.55
3	4.12	3	5.90	Dieldrin	4.40	Xenograde	12.08
4	4.18	4	6.57	Endrin	7.75	Prison System	6.80
5	5.15	5	5.88	Parathion	5.73	Minnesota Theater	8.02
6	4.57	6	4.68	Malathion	5.67	DDT	11.23
Total	27.87		33.81		38.06		54.54

[a] The segments in the continuous-related and continuous-scrambled segments are identical in this table. The segments were placed in the 4, 2, 5, 1, 6 and 3 positions during presentation in the latter treatment. The DDT passages in the last two columns were identical.

The analysis of two forms of the immediate true-false test yielded no significant difference due to form of the test so all subsequent analyses were based upon pooling of subjects across this factor. The mixed analysis of variance of the data from the immediate and delayed recall tests involved the four treatments related to thematic organization of the material and the two note-taking treatments as between-subjects variables. The immediate and delayed tests were employed as the within-subjects variable. This analysis yielded $F\ (1,192) = 6.29$, $p < .01$ for the effect due to note taking; the no-notes group achieved an average score of 51.5 correct items while the group that took notes achieved a mean score of 53.2 correct items. Of particular interest is the finding that, as in Experiment I, there were no significant differences ($p > .10$) among the groups on the immediate test. Thus, the means were 23.6 for the DRT theme, 22.4 for the DUR theme, 22.5 for the CRT theme, and 22.8 for the CST theme. However, the effect due to the interaction between thematic organization and delay of test yielded $F\ (3,192) = 5.00$, $p < .002$. Inasmuch as the effects on the immediate test were not significant this interaction reflects the significance of the differences among thematic organizations of the material on the delayed test. The order of the means for these groups differed from the order of those obtained in the previous analyses, they were: $\bar{X} = 85.6$ for the DRT group; $\bar{X} = 82.2$ for the CST group; $\bar{X} = 80.0$ for the DUR group; and $\bar{X} = 79.6$ for the CRT group. Calculations of t-tests indicate that of the latter three means are significantly ($p < .01$) different from the mean for the DRT group but are not significantly ($p > .05$) different from each other.

While there were slight differences in results from the two experiments the poorer performance under continuous related thematic material was reliably replicated. The differences in results from the two experiments may be due, in part, to the longer time allowed the subjects in Experiment II for the free-recall test.

Individual differences. The relationships between the individual difference variable (memory span) and performance measures are presented in Table 3. It can be seen in that table that only two of these correlations are significant (p = .05): one was the correlation between memory span and number of ideas produced (r = .38) for subjects in the DRT, note-taking treatment; the other was the correlation between memory span and the delayed true-false test (r = .42) in the CRT-note-taking treatment. These, of course, may have occurred by chance, and any conclusions can only be suggestive. Nevertheless, there is some consistency in these data since both memory-span scores (one based on number of digits recalled and the other based on correct recall of all digits in a triplet in the correct order) yielded similar correlations, all were positive, and all were found in the note-taking groups. By contrast the set of correlations between the immediate and delayed true-false scores and the memory-span tests, for the CRT no-notes group are all negative and although not significantly different from zero are significantly different from the aforementioned correlations. Nevertheless, when tests of parallelism for regression was made on these data, no significant F ratios were found, indicating that there was insufficient evidence for rejecting the hypothesis that the regression lines were parallel in the population.

Table 3

Summary of Correlations Between A Measure of Memory Span and Several Performance Measures

Based on Groups (n = 25) Receiving Variations in Thematic Relatedness and Note-taking Treatments

Experimental Treatment	Correlation Between: Immediate True-False Test and MS I[a]	Immediate True-False Test and MS II[b]	Delayed True-False Test and MS I[a]	Delayed True-False Test and MS II[b]	No. of Words Produced and MS I[a]	No. of Words Produced and MS II[b]	No. of Ideas Produced and MS I[a]	No. of Ideas Produced and MS II[b]
Discontinuous-Related Theme								
No Notes	-.01	-.04	-.13	-.23	.16	.14	.04	.00
Notes	.05	.10	-.04	-.07	-.03	-.04	.38*	.35*
Discontinuous-Unrelated Theme								
No Notes	-.02	-.09	.18	.15	.23	.19	.26	.12
Notes	-.10	-.13	.16	.13	.33	.12	.13	.07
Continuous-Related Theme								
No Notes	.08	.06	.04	.04	.03	.03	-.01	-.01
Notes	.01	.21	.26*	.42*	-.29	-.02	-.24	.04
Continuous-Scrambled Related Theme								
No Notes	-.27**	-.29***	-.31**	-.24**	.01	.00	.01	-.12
Notes	-.11	.04	-.15	-.18	-.02	.19	-.21	-.03

* Asterisked figures indicate differences between note-taking and no-notes groups in relationships between the individual difference variable and performance that may warrant further analysis or more explicit investigation to identify trait-treatment interactions. Nevertheless, only two of these are significant (a correlation of .38 is required for $p = .05$, two-tailed test).

** These are not significant, but are noted since they are opposite in sign from the single-asterisked items and occur for the no-notes group.

[a] Memory Span I. There were 12 CCC's in the task. For this score, a subject received a score of 1 if he had correctly recalled all 3 C's in their proper order. He received a score of 0 if any consonant was missing or if any consonant was misplaced. (Total possible score = 12).

[b] Memory Span II. The total possible points for this score were 72 (6 for each CCC). The subject received 1 point for each consonant recalled and 1 point for the correct positioning of each consonant in the CCC.

Discussion

The results of the present study indicate that note-taking and thematic relatedness were related to the ability of the subjects to recall the material to which they listened. These effects were especially to be noted on the free-recall task immediately after listening and on the delayed true-false test. This finding replicates one obtained by Crawford (1925) who also concluded "A careful analysis of all the facts seems to justify the conclusion that the immediate value of notes is less than the delayed-review value. This immediate value is of sufficient importance, however, to justify the practice of taking notes, even if there is no opportunity to use them later." (p. 384). There were no significant effects on either recognition and recall immediately after listening or on recognition a week later, due to variations in the listening-review intervals.

The results concerning note-taking clearly indicate the importance of this learning device as an aid to recalling material to which one listens. We share Crawford's (1925a) conclusion that "... taking notes on a point does not guarantee its being recalled at the time of the quiz but failing to take note of it very greatly decreases its chances of being recalled" (p. 289). Nevertheless, the reason for the beneficial effects of note-taking is not answered by any of the studies, either the present one or any of those listed in the references. If note-taking is beneficial because it serves as an external storage mechanism then it should interact with a review period, which it does not either in the present study or in an earlier one (Di Vesta & Gray, 1971, in press). If it serves as a coding mechanism, then it can be assumed that note-taking ought to be more beneficial with unrelated material than with related material which already has a built in organization. Our investigation of two variations of review periods points toward discarding the notion that note-taking serves only as an external storage mechanism. Nevertheless, other investigations in which internal relatedness of material is varied on other bases than those employed in the present study must be conducted before the view that note-taking serves as a coding mechanism is abandoned.

The fact that thematic relatedness, as a variable, was related to recall was not surprising. The surprising result was, of course, that the degree of retention was greatest, in both experiments, for the material that was not thematically related. This conclusion holds for the simple comparison of the retention of passages that were placed in a logical connected order with the retention of the same passages when the order of its paragraphs were scrambled. The result was even stronger where the paragraphs comprising the listening passage cover clearly different topics.

One might argue that the thematic-connectedness of the passages in the latter comparison is confounded with difficulty level. The final answer to this argument can only be determined by conducting an experiment in which difficulty is controlled. At present, there is no reliable procedure, of which the present authors are aware, for equating difficulty of connected discourse passages. However, on casual inspection it can be said that the passages in all conditions appeared to be of equal difficulty and in the discontinuous-unrelated thematic treatment one of the six paragraphs was identical to that presented in the discontinuous-related thematic condition. Recall in the latter condition was far superior to recall in the former condition. Certainly, the argument related to relative difficulty of passages can not hold for the treatment in which the paragraphs in the continuous related thematic material were scrambled and where recall of all but one of the passages was better in the continuous-related-scrambled organization than it was in the continuous-related organization.

Lacking contradictory evidence it shall be assumed for purposes of discussion that these results reflect the effect of variations in thematic relatedness. If this is the case, then the findings suggest a phenomenon somewhat akin to release from proactive inhibition (Wickens, 1970). Thus, where the material contains a large number of ideas which are all highly related (as they were in the passages describing the fire ant, its activities and control) then initial ideas can have a detrimental effect on the recall of subsequent ideas. On the other hand, where a passage is discontinuous in the sense that its constituent paragraphs cover different topics, then any proactive inhibition developed in a given passage is eliminated by the abrupt change in ideas. All this is speculative, of course; but it appears not only to offer a possible explanation of the results but also to suggest a means of testing the notion of "release from PI" with connected discourse.

These conclusions gain strong support from the findings reported by Anderson and Carter (1971) in a study conducted quite independently of the present one. They found that "The groups which learned verbatim, paraphrase, or verbatim-arranged sentences forgot significantly more than the group which learned new sentences" (quote from the abstract of the Anderson and Carter report). Their discussion of these results was as follows: "Such interference could not have occurred had the subjects rotely learned perceptual or acoustic features, since a paraphrase and its base do not share these surface features. The interference must have been at a semantic level. There is now a convincing case that interference theory accounts for the forgetting of meaningfully-learned connected discourse" (p. 6).

It is not clear why variations in listening-recall intervals did not produce an effect. In principle, variations in the length of the listening period are comparable to variations in massed and distributed practice. Similarly, frequency of review intervals is comparable to increases in opportunity for consolidation. In either case, it would be expected that at least one of the present treatments would have been clearly superior to another. Obviously, none was.

There are two possible reasons for the failure of the manipulation of listening-recall intervals to affect recall. First, the opportunity to review notes or to contemplate the material may have eliminated any detrimental effects that accrue from massed practice for passages that were no longer than 30-min. In this instance, another condition might have been provided in which the subjects were prevented from reviewing their notes or from considering the material during the consolidation period by performing another unrelated task. Second, the taking of notes may be itself a sufficient condition for consolidation of material learned. But in this instance, the test of the hypothesis would have required a condition in which the subject was not permitted to take notes, a condition included in the present experiment. While the former treatment suggested was employed in the previous study (Di Vesta and

Gray, 1971, in press) they were not crossed with the length of the listening-review period, as they must be for adequate test of these additional hypotheses.

In his early study conducted nearly half a century ago, Crawford (1925) pointed out the importance of investigating the dynamics of note-taking as follows: "... if note-taking is beneficial, as we have found it to be, there is great need of an analysis of note-taking practice and procedure in greater detail to teach students how to get the best possible results from their efforts" (p. 291). For a number of reasons the topic was not taken seriously until investigators such as M. J. A. Howe (1970) and Berliner (1970) only recently took the initiative. It appears that further analyses of the process as an instrumental activity in learning should be a productive source of hypotheses for investigations in the applied psychology of learning and instruction.

We have felt it important in this discussion to depart from the typical interpretation of data in order to suggest additional avenues for research on the topic of note-taking and listening. As M. J. A. Howe (1970) suggests ". . . more detailed investigation making use of notes made by students as a way of attempting to examine the relationships between learning and individual coding processes would be justified. For instance, it would be interesting to explore individual differences in note-taking and to examine the effects of different note-taking strategies upon subsequent recall of meaningful materials" (p. 63). Of particular importance, in this regard, are questions related to the entire issue of coding. For example, more needs to be known about the specific points made in notes and their relationship to what is recalled and what is not recalled. Furthermore, it is probably true that some students record notes in a verbatim way while others record notes in elaborate organizational schemes with a great deal of translation. Still others record notes verbatim but then reorganize, translate, and elaborate on them after the lecture. These are variations in encoding mechanisms. Investigations on such variables should provide considerable information on the cognitive processes.

References

Anderson, R. C., & Carter, J. Retroactive inhibition of meaningfully learned sentences. Urbana, Illinois: University of Illinois, 1971. (Mimeo)

Ausubel, D. P. Educational psychology: A cognitive view. New York: Holt, Rinehart and Winston, 1968.

Berliner, D. C. Effects of test-like events and note-taking on learning from lecture instruction. Amherst, Mass.: Technical Report, Center for Educational Research, March, 1970.

Carson, R. L. Silent spring. Boston: Houghton-Mifflin, 1962.

Crawford, C. C. The correlation between college lecture notes and quiz papers. Journal of Educational Research, 1925, 12, 282-291. (a)

Crawford, C. C. Some experimental studies of the results of college note-taking. Journal of Educational Research, 1925, 12, 379-386. (b)

Di Vesta, F. J., & Gray, G. S. Listening and note-taking. Journal of Educational Psychology, 1971, in press.

Eisner, S., & Rohde, K. Note-taking during or after the lecture. Journal of Educational Psychology, 1959, 50, 301-304.

Hebb, D. O. A textbook of psychology. Philadelphia: W. B. Saunders, 1966.

Howe, M. A. Introduction to human memory: A psychological approach. New York: Harper and Row, 1970.

Howe, M. J. A. Using students' notes to examine the role of the individual learner in acquiring meaningful subject matter. Journal of Educational Research, 1970, 64, 61-63.

Loveland, F. Classification in the prison system. In N. Johnston, L. Savitz, and M. E. Wolfgang (Eds.) The Sociology of Punishment and Correction. New York: Wiley, 1962.

Merrill, M. D. Transfer effects wtihin a hierarchical learning task as a function of review and correction on successive parts. [Office of Naval Research Technical Report No. 5, Training Research Laboratory, University of Illinois, Contract Nonr 3985 (04), September, 1964] Ann Arbor, Mich.: University Microfilms, 1965. No. 23-029.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. Journal of Experimental Psychology, 1959, 58, 193-198.

Watkins, J. G. Hypnosis. In N. L. Faberow (Ed.) Taboo topics. New York: Atherton, 1966. Pp. 96-106.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. Psychological Review, 1970, 77, 1-15.

Summary

The Effects of Congruence Between Passage Organization and an Imposed Strategy on Clustering and Recall of Textual Materials

Francis J. Di Vesta, Charles B. Schultz, and Timothy R. Dangel

Technical Problem

Simple statements typically are comprised of a concept name and a concept attribute, e.g., "Atweena (concept name) is a mountainous country (coucept attribute)." Statements such as this can be grouped into paragraphs in which the common organizational element is a concept name or a concept attribute, or in which no systematic organizational pattern is used. As they study written passages, learners may subjectively cluster statements in a similar way. The primary purpose of the present experiment was to compare the imposition of a subordinate strategy which was incongruent with the passage organization to a condition in which both passage organization and clustering strategy are congruent.

General Methodology

The subjects had four trials to learn a written passage about six imaginary nations. For the first three trials, subjects were given a three-minute study period and a six-minute writing period (i.e., the free-recall test). In one treatment, passages were organized by concept name, concept attribute, or by a random selection of statements. One third of the subjects in each passage organization group were instructed to use a name clustering strategy, one-third an attribute strategy, and the remainder were given no particular organizational instructions. On each trial, measures of statements correctly recalled, name clustering ratios, and attribute clustering ratios were obtained. In addition, a secondary organization score was obtained reflecting the consistency of internal organization of clusters from one cluster to the next.

Technical Results

The subjects who studied passages organized by either name or attribute recalled more than subjects who studied passage without an organizational pattern. Preference was also found for the name clustering strategy in contrast to the attribute strategy. The adoption of the less-preferred attribute strategy was gradual when the subject studied the attribute passage and was free to choose his own clustering strategy compared to a condition which was similar in all respects except that the attribute strategy was imposed. One of the most interesting findings was that incongruence between passage organization and clustering strategy resulted in more recall than congruence.

Educational Implications

These findings underline the importance of making the organizational pattern of learning materials apparent to the learner. The learner must concentrate on defining an organizational pattern for loose and unstructured textual materials if recall of the material is not to be impaired. Moreover, the more organizational cues the instructor provides for the learners, the greater the recall. Since often textual material contains both names and attributes, the characteristics of both dimensions should be brought to the learner's attention.

Implications for Further Research

The present study can be extended in at least two directions. One tact is to examine the effect of the matrix clustering strategy, implied by the use of a system of secondary organization in the present study, on recall and on other processes such as inference-making. In addition, the learner's clustering strategies and their relation to recall could be examined when the instructional materials are presented orally (i.e., approximating a lecture form) rather than in textual form.

The Effects of Congruence Between Passage Organization and an Imposed Strategy on Clustering and Recall of Textual Materials

Francis J. Di Vesta, Charles B. Schultz, and Timothy R. Dangel

Textual materials can be organized into paragraphs in which the central organizing elements is a concept name or a concept attribute, or the statements can be placed in paragraphs which have little or no apparent organizational structure. For example, the sentence, "Atweena (concept name) is an island nation (concept attribute)," can be placed in a paragraph with other statements about Atweena (e.g., "In Atweena, the society is largely industrial"), in a paragraph with statements about the geography of different nations (e.g., "A mountainous terrain characterizes much of Egrama"), or in a paragraph with randomly selected statements. Just as passages may be organized according to concept name or attribute, learners themselves may <u>subjectively</u> cluster statements in a passage using similar rules in an effort to facilitate learning and improve retention.

In earlier studies, Frase (1969) and Schultz & Di Vesta (1972) have demonstrated that passages organized by name or attribute result in greater recall than passages without an organizational pattern, that clustering in recall is influenced by passage organization, and that the concept name clustering strategy is preferred over the concept attribute strategy. Since clerical employees and high school students, respectively, were used as subjects in these experiments, the present investigation provides a basis for generalizing these findings to other populations, college students, who presumably are more adept verbally and who have had more experience studying textual materials with different organizational patterns, were employed.

A particularly interesting finding from the Schultz & Di Vesta (1972) study was that name and attribute clustering strategies were adopted at different rates. Thus, concept name clustering strategy was found to be used in the first encounter with the passage organized by name, but the concept attribute strategy was adopted only after several exposures to the passage organized by attribute. The data imply that when the dominant clustering strategy is congruent with the passage organization, it is immediately adopted; when the subordinate strategy is congruent with the passage, it is gradually adopted perhaps because of interference from the dominant strategy. This implication, stated as an hypothesis, was tested directly in the experimental design of the present study. Accordingly, subjects were <u>required</u> to adopt either the dominant strategy, the subordinate strategy or, in a third treatment, were free to adopt any strategy they wished to use.

Finally, Schultz & Di Vesta (1972) found that recall was impaired and errors more frequent during the first trial of learning a passage organized by attribute when reliance on the incongruent name strategy was presumably greatest. This finding implies that a clustering strategy which is incongruent with the organizational structure of the passage may inhibit learning and retention of textual materials. Accordingly, it was hypothesized, for the present study, that the imposition of a clustering strategy which was congruent with the passage organization would result in greater recall and fewer errors than when the imposed passage and clustering strategy were incongruent with the passage organization.

Method

Design

College students were given three trials to learn a passage which described six imaginary nations. Each trial was comprised of a brief study period followed by a free-recall test. After the third trial, subjects completed a task which was intended to prevent rehearsal of the passage before the administration of a fourth free-recall test. The number of correct responses and clustering ratios were obtained on each free-recall test.

Three levels of passage organization (concept name, concept attribute, and random sentence sequence) were orthogonally crossed with three sets of instructions for grouping the statements in the passage (name grouping, attribute grouping, and no specific grouping instructions). These manipulations imply a 3 x 3 x 4 mixed analysis of variance design, with repeated measures on the last factor, the free-recall test.

Subjects

The 99 subjects were undergraduate students enrolled in an introductory educational psychology course at The Pennsylvania State University. Students were awarded standard score points toward their grade for participation in the experiment. The experimental sessions were conducted with groups of nine subjects, each of whom was randomly assigned to a different experimental condition.

Experimental Materials

The study passage was similar to that used in the earlier experiment (Schultz & Di Vesta, 1972). It consisted of 36 statements describing six imaginary nations called Atweena, Brontus, Egrama, Nurovia, Bismania, and Galbion. The geography, government, mood, technology, population growth, and birth rate were described for each nation. It was possible, therefore, to use the same 36 statements to construct three passages with different organizational patterns. In one the statements were arranged into six paragraphs each with six sentences about the same nation. In a second organizational pattern, each paragraph was comprised of statements about the same attribute. A third passage contained six paragraphs with statements randomly selected from the same pool of statements as that used in the preceding conditions.

Procedure

The experiment was conducted in a language laboratory. Each of the nine subjects worked in an isolated booth and wore earphones. Only the timing signals and instructions for all phases of the experiment were tape-recorded and received by the subject as he read a printed version of the instructions included in an answer booklet. The language laboratory facility permitted the simultaneous presentation of instructions for nine different experimental conditions.

The subject's task was to learn a passage containing descriptions of six imaginary nations. Each study-recall trial consisted of a 3-min. study period, during which time subjects were instructed to read and study the complete passage without the assistance of notes, and a 6-min. writing period, during which time they were instructed to write down all the statements they could remember (i.e., the free-recall test).

Organization treatments. One experimental variation consisted of manipulating the organization of the passage as described in the Materials section. Briefly, some subjects read a passage with paragraphs organized by name, (Concept Name Passage), while others studied a passage with paragraphs organized by attribute (Concept Attribute Passage) or a passage comprised of randomly ordered statements (Random Order Passage).

Instructional sets. One-third of the subjects assigned to each of the organization treatments were instructed to learn the passage by grouping the material according to the concept name (Concept Name Instructions). Another one-third of the subjects assigned to each of the organization treatments were given instructions to learn the passage by grouping the material by concept attribute (Concept Attribute Instructions).

The remaining third of the subjects assigned to each of the organization treatments were instructed to either rehearse the statements, to group them, to use a mnemonic "gimmick," or to use whatever plan or strategy they felt would help them remember the passage (No Organizational Instructions).

Intervening task. Following the third trial, subjects were given a ten-minute task designed to prevent rehearsal of the passage and to provide a measure of memory decay. The intervening task was based on procedures used in the Peterson & Peterson (1959) study of short-term memory. The subjects first heard a three-consonant syllable which was immediately followed by a three-digit number. The number served as a reference point from which subjects counted backwards aloud by threes (or fours) in time with a signal presented at half-second intervals before attempting to recall the syllable. There were 12 such trials administered. For the first four trials, the time spent counting backwards was three seconds; for the second four trials, the time was six seconds; and for the last four trials, subjects counted for nine seconds. Immediately following the short-term memory task, a fourth free-recall test was administered without a study period.

Association test. An association test followed the fourth writing period. For this task, 12 cue statements from the passage were read to the subjects. They were instructed to respond to each statement by writing the first statement from the paragraph that came to mind. For example, when the cue, "Egrama is characterized by a mountainous terrain," was read, the subject could respond with another statement about Egrama (suggesting name organization) or with a statement about the geography of a different country (suggesting attribute organization). Since each cue statement contained a name and an attribute, subject's responses could be analyzed to determine the ratio of name or attribute associations he made. A score based on the association test was computed by subtracting attribute ratio from name ratio and adding 1.00. This procedure yielded a range of zero to two. Low scores reflect attribute clustering while high scores reflect name clustering.

A post-experimental questionnaire was administered at the conclusion of the experiment. The subject was asked to describe his strategy for learning the statements in the passage. The subject rated the instructions for grouping the statements on a scale consisting of the following points: very helpful, somewhat helpful, neither helpful nor interfering, somewhat interfering, or very interfering. The subject also listed his credits in history and the social sciences.

Scoring

The free-recall protocols were scored according to procedures described more fully by Schultz & Di Vesta (1972). They consisted of obtaining measures of correct statements recalled, name clustering ratios (which reflect the amount of organization in free-recall by concept name) and attribute clustering ratios (which reflect the amount of organization in free-recall by concept attribute).

For scoring purposes a "cluster" was defined as two or more consecutive statements about the same name or about the same attribute. Another score was also obtained, to reflect the degree of consistency of organization across clusters for a given trial. It showed whether the sequence of statements within the dominant bases for organization (i.e., name or attribute) was in the same inter-cluster order. Thus, for example, if the dominant basis for organization was name, the statements would be organized within each name by attributes. With perfect consistency (represented by a score of 1.0) the attributes would be recalled in exactly the same order from one cluster to the next. With complete inconsistency represented by a score of zero, the order of attributes recalled in one cluster would bear no relationship to the order of attributes recalled in the subsequent clusters.

Results*

Analyses were conducted of clustering scores and statements correctly recalled, on each of the four free-recall trials. Additional analyses were made of the association scores and the sequence within clusters (SWC) scores for the fourth trial. Newman-Keuls procedures were used for all multiple comparisons ($p < .05$). Because of heterogeneity of variance in

*All analyses of variance referred to in this section are summarized in Table 2.

the analyses of repeated measures, corrections were made for analyses of independent (nested) factors with heterogeneous variances and for analyses based on repeated measures.

Induction of Treatments

The association test was used to determine the extent to which the experimental conditions were induced. High scores (i.e., up to 2.0) reflect name organization and low scores (i.e., to 0.0) reflect attribute organization. An analysis of variance of these data yielded F (2,90) = 9.35, $p < .01$ for the effect due to passage organization in which the following order of means was obtained: $\bar{X} = 1.55$ for the Concept Name Passage (CNP) organization, $\bar{X} = 1.26$ for the Random Order Passage (ROP) organization, and $\bar{X} = .87$ for the Concept Attribute Passage (CAP) organization. The ROP and CNP organizations differed from the CAP organization ($p < .05$), but not from each other. The analysis of variance of free association scores yielded F (2,90) = 3.53, $p < .05$ for the effect due to instructions. Mean scores for the Concept Name Instructions (CNI), No Organization Instructions (NOI), and Concept Attribute Instructions (CAI) groups were $\bar{X} = 1.47$, $\bar{X} = 1.08$, and $\bar{X} = 1.14$, respectively. The scores for the CNI group were greater than those of the CAI and NOI groups ($p < .05$) which did not differ from each other. Thus, in both the passage organization and instructional set treatments, the scores of the CN and CA groups differed from each other on the association test. These results implied that the type of organization (name or attribute) employed by the subject was influenced or induced by cues from both the passage and the instructions.

Clustering

Separate analyses of variance were made on the amount of name clustering and attribute clustering during recall. An additional analysis was made of scores which reflect the consistency of the sequence of the order of statements within clusters (SWC). A summary of the means based on the clustering scores is presented in Table 1.

Name clustering. The analysis of variance of name clustering ratios yielded $\underline{F}$ (2,90) = 6.66, $\underline{p} < .01$ for the effect due to passage organization. The mean name clustering scores for the three passage organizations were $\bar{X}$ = 68.70 for the CNP organization, $\bar{X}$ = 48.75 for the CAP organization, and $\bar{X}$ = 45.63 for the ROP organization. According to the Newman-Keuls tests, the name clustering ratio for the CNP organization was higher than those for the CAP and ROP organizations which did not differ from each other. The analysis of variance of name clustering scores yielded $\underline{F}$ (2,90) = 35.20, $\underline{p} < .01$ for the effect due to instructions. The mean scores for CNI, CAI, and the NOI conditions were $\bar{X}$ = 85.23, $\bar{X}$ = 49.54, and $\bar{X}$ = 28.30 respectively. The name clustering ratios for each group were significantly different from each of the other groups ($\underline{p} < .05$).

Attribute clustering. The analysis of the attribute clustering ratio yielded $\underline{F}$ (2,90) = 5.01, $\underline{p} < .01$ for the effect due to passage organization in which the ratios of the CAP organization ($\bar{X}$ = 49.92) and the ROP organization ($\bar{X}$ = 31.80) were higher than those of the CNP organization ($\bar{X}$ = 22.56). The CAP and ROP organizations did not differ from each other. A similar analysis of the attribute clustering ratio for the effect due to instructions yielded $\underline{F}$ (2,90) = 39.19, $\underline{p} < .01$ in which each instructional group is different from the others. The means

Table 1

Summary of Mean Scores on Each Dependent Variable

For Each of the Experimental Conditions

	Organization of Passage			
Instructions	Random	Attribute	Name	Total
	Correct Statements			
None	11.41	12.07	11.57	11.68
Attribute	9.82	12.41	17.48	13.23
Name	9.80	14.91	13.25	12.65
Total	10.34	13.12	14.10	
	Name Clustering Ratios			
None	41.91	42.18	64.52	49.54
Attribute	29.52	13.45	41.93	28.30
Name	65.45	90.61	99.63	85.23
Total	45.63	48.75	68.70	
	Attribute Clustering Ratios			
None	25.95	43.68	15.68	28.44
Attribute	56.02	81.20	50.98	62.73
Name	13.41	3.89	.98	6.09
Total	31.80	42.92	22.54	
	Inter-Cluster Consistency (Fourth Trial Only)			
None	31.27	59.09	48.82	46.39
Attribute	50.36	67.00	91.36	69.58
Name	47.91	81.45	89.23	72.70
Total	43.18	69.18	76.30	
	Ratings of Instructions			
None	2.91	2.55	2.73	2.77
Attribute	2.64	2.55	1.82	2.33
Name	2.36	2.00	2.45	2.27
Total	2.63	2.36	2.33	

Table 2

Summary of Analyses of Variance for Each of the Dependent Variables Employed

		Correct Statements		Name Clustering		Attribute Clustering		Sequence Within Clustering (4th Trial)	
Source	df	MS	F	MS	F	MS	F	MS	F
Between Subjects									
Organization	2	502.30	5.94***	20674.80	6.66***	13743.60	5.01***	10030.50	14.79***
Instructions	2	81.23	.96	109261.00	35.20***	107452.00	39.19***	6814.43	10.05***
Org. x Instr.	4	235.07	2.78**	4756.35	1.53	4240.96	1.55	1019.17	1.50
Error	90	84.50		3103.64		2742.09		678.06	
Within Subjects									
Trials	3	1991.85	204.51***	567.00	1.34	1458.96	4.77***		
Org. x Trials	6	24.57	2.53**	154.84	.37	365.58	1.19		
Instr. x Trials	6	17.52	1.80*	790.27	1.87*	471.62	1.54		
Org. x Instr. x Trials	12	11.25	1.16	467.77	1.10	635.02	2.08***		
Error	270	9.74		423.59		306.06			

*$p < .10$

**$p < .05$

***$p < .01$

of these groups were: $\bar{X}$ = 62.74 for CAI, $\bar{X}$ = 28.44 for NOI, and $\bar{X}$ = 6.10 for CNI. The analysis of the effect due to trials yielded F (3,270) = 4.77, $p < .01$. Mean scores on Trial 1 ($\bar{X}$ = 27.27) differed from Trial 3 ($\bar{X}$ = 33.99) and Trial 4 ($\bar{X}$ = 36.31) but not from Trial 2 ($\bar{X}$ = 32.11).

The analysis of attribute clustering ratios yielded a three-way interaction (F [12,270] = 2.08, $p < .05$) for Instructions, Organization and Trials. None of the two-way interactions were found to be significant ($p > .10$). In order to examine the triple interaction in more detail, an analysis of simple effects for the Trials x Instruction interaction was conducted at each level of passage organization. At the CNP and ROP organization levels, there were no Trial x Instructions interactions. However, a Trial x Instructions interaction was obtained at the CAP organization level (F [6,90] = 3.67, $p < .05$). As can be seen in Figure 1, subjects in the CAI and CNI conditions do not change in their use of the attribute clustering strategy across trials. However, according to the Newman-Keuls test, attribute clustering scores in Trials 1 and 2 differed from those on Trials 3 and 4 for subjects in the NOI group. On Trial 1 attribute clustering scores for the CA Instructions group were higher than those of the NOI and CNI groups which did not differ from each other. By Trial 4, attribute ratios of both the CAI and NOI groups were higher than those of the CNI group, but not different from each other. This finding clearly supports the hypothesis that the adoption of a subordinate clustering strategy is gradual, occurring only after several encounters with a passage characterized by a given organization (i.e., name or attribute).

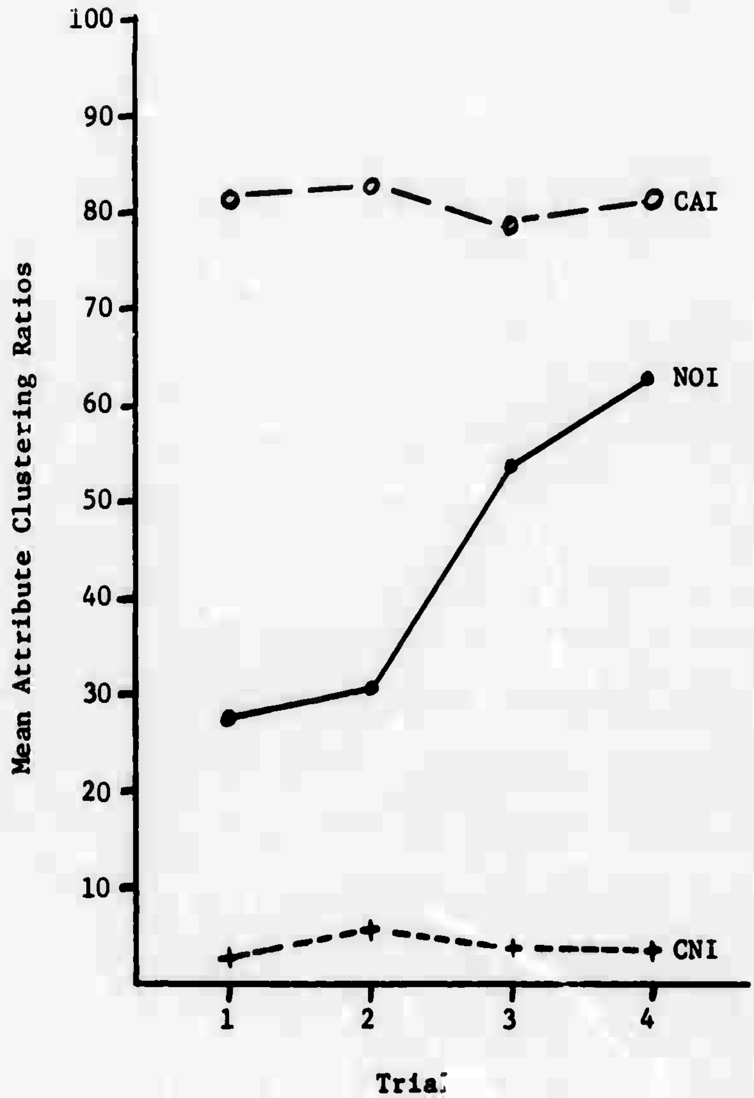

Figure 1. Attribute clustering scores of the concept attribute organization group for three instruction groups across trials.

It was also hypothesized that unlike the gradual adoption of the concept attribute clustering strategy when no organization instructions were provided, the concept name clustering strategy would be adopted in the first trial. In order to test the interaction implied by this hypothesis, a clustering score was required which would reflect both concept name and attribute factors. Accordingly, a combined clustering score was obtained as follows: concept name ratio - concept attribute ratio + 100. High scores (up to 200) imply name organization and low scores (to 0.0) imply attribute clustering.

A contrast to test the interaction between name and attribute passage organization in Trial 1 and Trial 4 for subjects in the NO Instruction group yielded $\underline{t}$ (270) = 2.67, $\underline{p} < .01$. The means for the above analysis are graphically displayed in Figure 2 where it may be seen that the clustering data clearly imply that the strategy of organizing by name was immediately adopted and maintained, while the adoption of the clustering strategy of organizing by attribute was very gradual over trials. By Trial 4, the CNP organization differed significantly from the CAP organization ($\underline{t}$ [90] = 2.05, $\underline{p} < .05$). This finding replicates an earlier one obtained by Schultz & Di Vesta (1972) on the selection of dominant and subordinate strategies.

<u>Internal consistency of clusters</u>. The analysis made of the amount of sequencing within clusters (SWC) were obtained from Trial 4, the means for which are summarized on Table 1. The analysis of SWC's yielded $\underline{F}$ (2,90) = 14.79, $\underline{p} < .001$ for the effect due to organization. The direction of the mean SWC scores were as follows: $\bar{X}$ = 76.30 for CN organization, $\bar{X}$ = 69.18 for CAP organization, and $\bar{X}$ = 43.18 for ROP organization. The CAP and CNP organizations differed from the ROP

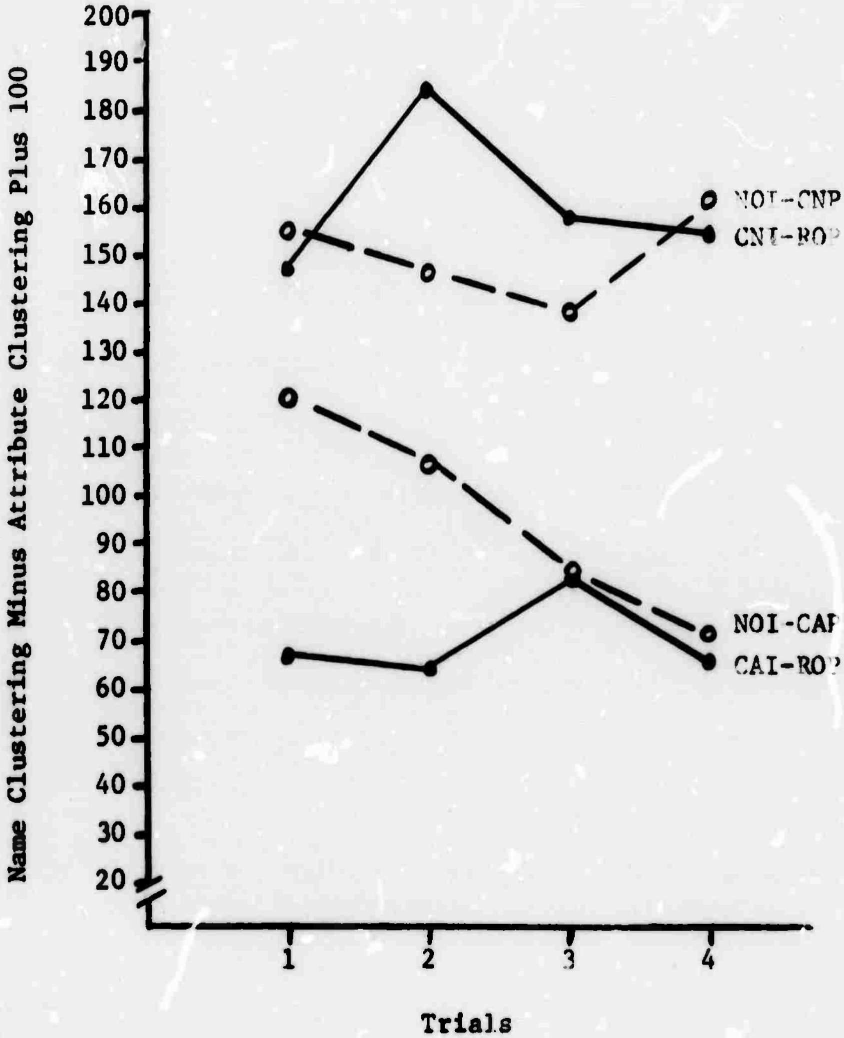

Figure 2. Name clustering score minus attribute clustering score plus 100 for four instruction and organization groups across trials.

organization, but not from each other. A similar analysis of the effect of instructions yielded $\underline{F}$ (2,90) = 10.05, $\underline{p} < .001$. Scores for the CAI ($\bar{X}$ = 72.70) and the CNI ($\bar{X}$ = 69.58) conditions were higher than for the NOI condition ($\bar{X}$ = 46.39) but not from each other.

In order to determine whether one strategy was more dominant than the other in the present study, a $\underline{t}$ test of correlated means of the name and attribute clustering ratios was made (McNemar, 1969, p. 113-114). This analysis yielded $\underline{t}$ (98) = 3.11, $\underline{p} < .01$, implying preference for name over attribute strategies, thereby providing direct support for earlier findings by Frase (1969) and Schultz & Di Vesta (1972).

<u>Recall of Correct Statements</u>

The analysis of variance of statements correctly recalled yielded $\underline{F} < 1.0$ for the effect due to instructions. A similar analysis of the effect due to passage organization yielded $\underline{F}$ (2,90) = 5.94, $\underline{p} < .01$ in which both CNP ($\bar{X}$ = 14.10) and CAP ($\bar{X}$ = 13.13) recalled more than the ROP group ($\bar{X}$ = 10.34) but did not differ from each other. A Passage Organization x Instructions interaction was also obtained ($\underline{F}$ [4,90] = 2.78, $\underline{p} < .05$) as represented in Figure 3. A comparison of simple effects with the Newman-Keuls test yielded the following: (a) recall by the NOI group did not differ across different passage organization treatments; (b) recall by the CNI group was greater in both the CNP and CAP treatments than in the ROP treatment; (c) recall by the CAI group was greater in the CNP treatment than in the CAP or ROP treatments; and (d) recall by the CAI group with the CNP treatment was greater than that of the CNI and NOI groups who also studied the CNP passage. Thus, the interaction effect appears to be attributable largely to the superior

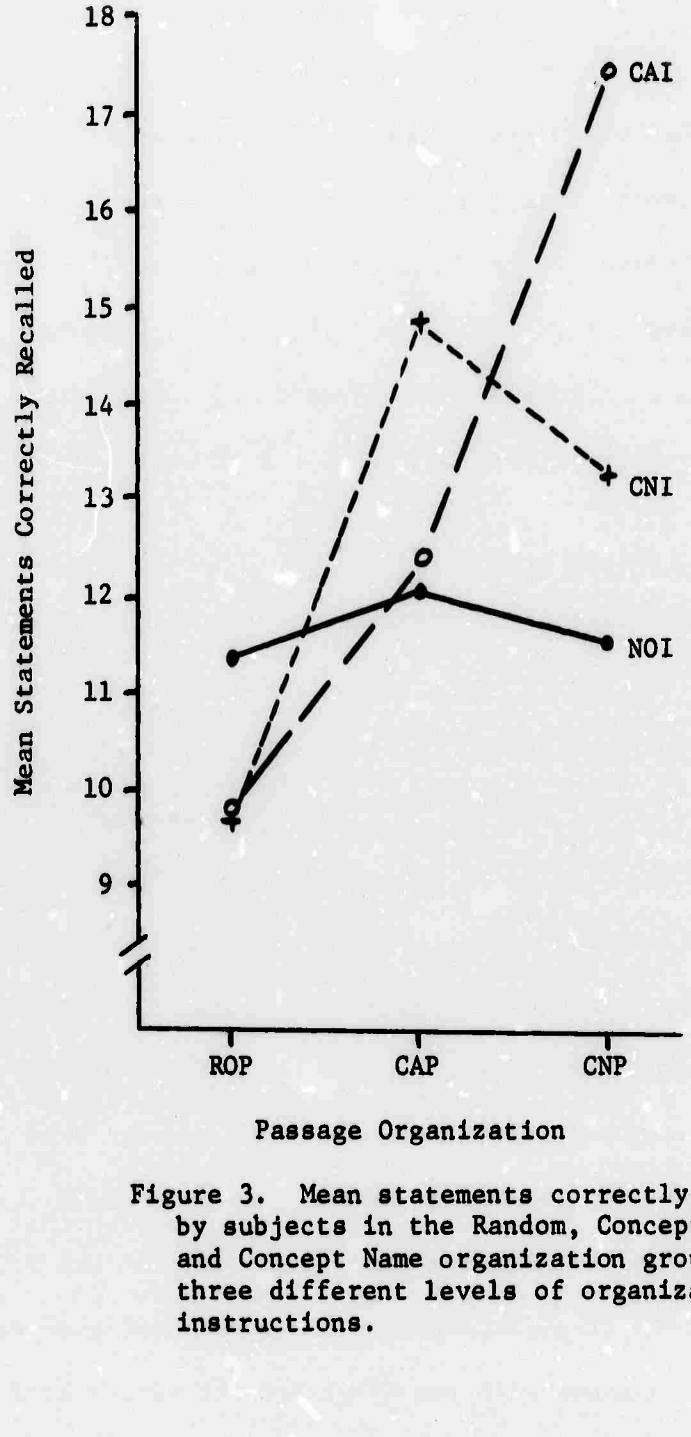

Figure 3. Mean statements correctly recalled by subjects in the Random, Concept Attribute, and Concept Name organization groups with three different levels of organizational instructions.

recall of subjects with concept name instructions who studied the concept attribute passage.

The analysis of variance of correct statements also yielded a significant effect, $\underline{F}$ (3,270) = 204.51, $\underline{p}$ < .01, due to trials, as well as a significant Trials x Organization interaction, $\underline{F}$ (6,270) = 2.52, $\underline{p}$ < .05). The most important features of this interaction, displayed in Figure 4, is the tendency for recall in the ROP group to become increasingly depressed relative to that of the CNP and CAP groups, with additional trials, until by Trial 4 it was significantly lower than the two groups receiving organized passages.

<u>Effect of Congruency Between Passage Organization and Instructional Sets</u>

A primary interest in the present study was the comparison of recall by subjects whose passage organization was either congruent or incongruent with the clustering instructions they received. It was hypothesized that congruence between passage and instructions would result in greater recall than incongruence. The means for statements correctly recalled under these conditions are summarized in Table 1. There it may be seen that the results are opposite to the hypothesized direction; thus, scores for groups in which both manipulations were congruent (i.e., attribute passage-attribute instructions and name passage-name instructions) were lower than scores for groups in which both manipulations were incongruent (i.e., attribute passage-name instructions and name passage-attribute instructions). A test of the interaction implied by the direction of these means yielded $\underline{t}$ (90) = 2.43, $\underline{p}$ < .02 indicating significantly greater recall when the passage organization and clustering strategies were incongruent than when they were congruent.

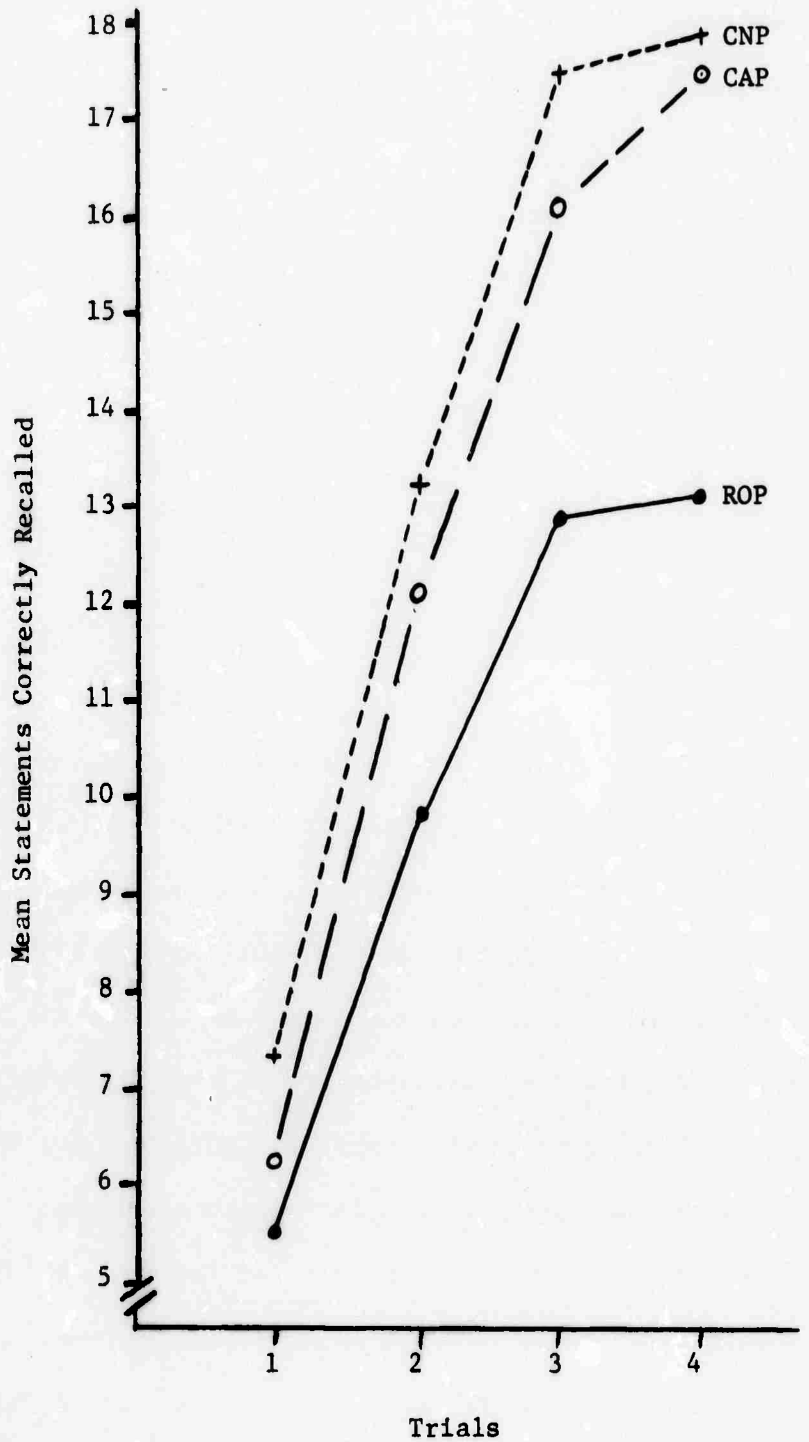

Figure 4. Mean statements correctly recalled for three different passage organizations across trials.

The results described in the aforegoing paragraph, and based upon recall scores, were supported by subjects' ratings of the helpfulness of the instructional treatments. These subjective reports provide an independent rating inasmuch as the subjects received no evaluational feedback, neither norm-referenced or criterion-referenced, about their performance. The analysis of particular interest is that concerned with the interaction between Name and Attribute Instructional Sets and Name and Attribute Passage Organization. The interaction contrast involving these four cells yielded $\underline{t}$ (90) = 2.18, $\underline{p} < .05 > .02$. As may be seen in Table 1, the magnitude of these ratings varied directly with the magnitude of the recall scores. Thus, when passage organization and instructional set were incongruent the instructions were rated as more helpful (represented by lower scores) than when two variables were congruent.

Discussion

Several findings from the present study support those reported earlier by Frase (1969) and Schultz & Di Vesta (1972). One of the fundamental results was that subjects who studied passages organized in some manner (i.e., either by name or attribute) were able to recall more information than those subjects who studied passages without an organizational pattern. Apparently, statements acquire more meaning when they are grouped into what may be called higher order memory units and, as a result are more readily accessible at the time of recall. The fact that more ideas were recalled from organized passages than from unorganized ones is compatible with the notion that organization is an important component of meaning or cognitive structure (Ausubel, 1968).

Higher recall in this instance of organized passages reflects the existence of larger memory units (see, e.g., Tulving & Patkau, 1962).

Another result in this study provided support for the dominance of the name clustering strategy during recall. As indicated by Schultz & Di Vesta, (1972) this preference for the name clustering strategy over the attribute clustering strategy may be due to the history of the learner whose primary experience has been with organization of attributes within broad conceptual categories (i.e., name) which act as subsumers. An alternative explanation, and perhaps more important one from the standpoint of instructional strategy, would suggest that there are fewer names and that their labels are easily discriminated. The attributes, although also six in number, have six values each. Accordingly, all attributes and their respective values comprise a larger (n = 36) set and are not readily discriminated or associated with the appropriate label (or name) without a great deal of experience on the part of the learner. It is apparent that in such learning, the name can be acquired in a one-stage process whereby one name is stored as a discriminable unit from another. Then when the attributes are to be associated with them, the attributes are subsumed as specific entities. However, when attributes are employed as organizing units, the individual who attempts to associate each entity with a name will have 36 independent statements to recall (e.g., Bismania is mountainous; Egrama is a flat plain; etc.). This procedure is certain to produce a great deal of memory strain unless the learner organizes the material into another set of higher order units such as geographical features, modes of transportation, and the like. But even this strategy requires an additional stage in learning compared to the strategy of learning by name first. It requires the learner to identify the distinctions among names, the structure of the set of attributes, and the values within each set. Because this strategy requires additional effort in the initial stages of learning it is less likely to be adopted immediately (or spontaneously) by the learner.

An especially unique feature of this study was the derivation of a score for intercluster consistency which, in effect, takes into account clustering strategies based on two levels of organization. It appears that such organization is especially important to prose (i.e., connected discourse). For example, the learner first organizes by name and then by attribute within name. Reflected in the data obtained from the subjects in this study, was the finding that there is consistency even in the subordinate levels of organization- that is, when the primary organization was by name, the subject in the later stages of learning also organized the secondary level to retain an order among attributes, from one name to the next. As shown in Table 1, the set of four cells in the present study which were characterized by the highest intercluster consistency were those that were cued by both passage organization and by instructions. The passage that was characterized by the least intercluster consistency was that in which cues were not provided by either treatment. The treatments in which cues were provided either by passage organization or by instructions (but not both) were in between these two extremes. Accordingly, the availability of the hypotheses regarding which strategies facilitate recall affect the degree of both primary and secondary organization.

Of special interest in the present study was the finding that incongruency between passage organization and instructional treatments facilitated recall of correct statements to a greater extent than did congruency. This finding did not support the initial hypothesis that instructions to employ a given strategy would summate with congruent passage organization to facilitate recall. Nevertheless, these findings can be clearly explained in terms of Restle's (1962) model. He contends that difficulties in cue learning are encountered when the subject employs strategies that conflict with the strategies intended by the experimenter. Such situations are encountered when the subject is given a passage organized according to attribute with no instructions other than those to memorize and, later, to recall the material. The subject then has difficulty because he must now abandon the dominant name strategy. Accordingly, the learning curve data, based on performance of the group that must learn the passage organized by attributes, reflects a gradual change from the strategy of organizing by name to the strategy of organizing by attribute. The gradual adoption of a strategy, where the correct strategy is not clear, differs radically from the rather abrupt and spontaneous adoption by the group instructed to use the appropriate (attribute) strategy. This is a clear replication of one reported earlier by Schultz & Di Vesta (1972). The statement by Johnson, Fishkin and Bourne (1966) that: "... instructions which include explicit labels for stimulus dimensions, which indicate only one dimension will be relevant and which demonstrate a possible solution to a problem all combine to induce a hypothesis-testing type of situation [p. 70]" seems as appropriate to the adoption of a strategy of selecting a basis for organizing connected discourse material as it is to concept learning.

However, the main thesis of Restle's (1962) model is supported in this study by the effect of the interaction between passage organization (i.e., name and attribute only) and instructional treatment (i.e., name and attribute only) on recall, where incongruency appeared to be more facilitative than congruency. Furthermore, it will be recalled that the subjects rated the incongruent conditions as more helpful than the conditions in which the instructional sets coincided with passage organization. These results can be explained on the basis that incongruent instructions tend to define, for the learner, the two fundamental rules for organizing the material to be learned. In fact, by employing these two rules he can form a matrix (and it is likely under the conditions of this particular experiment that he will do so) that would eventually permit him to make application (transfer) to a variety of situations. This conclusion is parallel with one by Hagen, Meacham, and Mesibov (1970) who indicate that "Verbal labels which are imposed externally are irrelevant, and even distracting for the individual who does not utilize them for task performance [p. 57]." In the present task, when a name organization is imposed, the attribute instructions provide supplementary information; similarly, for attribute organization with name instructions. Thus, the two rules are given, one implied in the organization and the other induced by instructions, and consistent with Hagan, Meacham, & Mesibov's (1970) comment is the implication that instructions, as presented in a task involving two dominant strategies, even though superficially incongruent with passage organization, need not be irrelevant unless they cannot be, or are chosen not to be used by the subject when performing the task.

Apparently, the subject found them useful since the data imply that he did use them and, in addition, he rated them as helpful. In actuality, "incongruent" instructions in a task involving two strategies are more informative than congruent ones inasmuch as they exhaust the space that must be scanned by the learner, thereby facilitating his performance as it did in the present experiment.

References

Ausubel, D. P. Educational psychology: A cognitive view. New York: Holt, Rinehart and Winston, 1968.

Frase, L. T. Paragraph organization of written materials: the influence of conceptual clustering upon the level and organization of recall. Journal of Educational Psychology, 1969, 60, 394-401.

Hagen, J. W., Meacham, J. A., & Mesibov, G. Verbal labeling, rehearsal, and short-term memory. Cognitive Psychology, 1970, 1, 47-58.

Johnson, P. J., Fishkin, A., & Bourne, L. E. Jr. Effects of procedural variables on reversal and interdimensional shift performance: Part II. Psychonomic Science, 1966, 4 (2), 69-70. Part I, 63-64.

McNemar, Q. Psychological statistics, New York: Wiley, 1969.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual items. Journal of Experimental Psychology, 1959, 58, 193-198.

Restle, F. The selection of strategies in cue learning. Psychological Review, 1962, 69, 329-343.

Schultz, C. B., & Di Vesta, F. J. The effects of passage organization and note-taking on the selection of clustering strategies and on recall of textual materials. Journal of Educational Psychology, 1971, in press.

Tulving, E., & Patkau, J. E. Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. Canadian Journal of Psychology, 1962, 16, 83-95.

Summary

Recitation Strategies II: The Effects of a Learner-Sustaining Climate and Encoding on Retention of Facilitators and Debilitators

Charles B. Schultz and Timothy R. Dangel

Technical Problem

Recitation has been found to be a stressful condition which improves retention for some learners (facilitators) and depresses learning for others (debilitators). One reason for the relatively low performance of debilitators may be that they disrupt short-term storage with task-irrelevant thoughts associated with anxiety-producing sitmuli. The debilitative effects of these intrusions may be reduced by minimizing the threat to the learner's self-esteem which is posed by the recitation situation and by requiring the learner to encode instructionally relevant information. It was expected that a learner-sustaining climate and translation of recitation answers would facilitate retention of debilitators compared to a directive climate and verbatim response mode.

General Methodology

Groups of six Ss recited answers to 18 recitation questions asked by E by finding the answers from printed textual material. The Ss were evenly divided between facilitators and debilitators according to the Achievement Anxiety Test. Climate was manipulated by minimizing the evaluative aspect of the recitation exercise for half of the Ss (sustaining) and emphasizing it for the others (directive). Finally,

PRECEDING PAGE BLANK

half the Ss were instructed to translate their answers in the printed material while the remainder were told to make verbatim responses. A 30-item multiple-choice test was given after the recitation period. This test included 18 items based on the answers to the recitation questions (intentional learning) and 12 items based on information included in the printed material but which Ss were not directed to examine (incidental learning). The State Anxiety Test was administered immediately after the recitation period and again after the multiple-choice test.

Technical Results

The sustaining climate was particularly effective in facilitating retention of information which was instructionally relevant, but which Ss were not directed to examine by the recitation questions. The translation response mode was especially effective in improving retention of the answers given to the recitation questions during the experimental "class." The retention scores of debilitators in the sustaining-translation condition were higher than of debilitators in the directive-verbatim condition.

Educational Implications

The present findings can be applied to instructional settings quite directly. Clearly, it is important for instructors to provide opportunities for learners to put the material they study into "their own words." The findings suggest impromptu responding and oral presentations from notes rather than the reading of reports prepared by the student or by others. This is particularly necessary when the learning occurs in moderately stressful conditions such as in the recitation exercise. The maintenance of a learner-sustaining climate facilitates incidental learning for all students, but is a condition which improves both incidental and intentional learning for debilitators.

Implications for Further Research

The present sutdy could be elaborated upon. Advanced organizers, labels, and other devices to facilitate encoding may also have facilitative effects on the retention of debilitators. Climate variables could also be manipulated. These variables include the use of direct and indirect categories of teacher verbal behavior which are important components of many instruments used in systematic observation of classroom verbal behavior.

Recitation Strategies II: The Effects of a Learner-Sustaining Climate and Encoding on Retention of Facilitators and Debilitators[1]

Charles B. Schultz and Timothy R. Dangel

A relatively large proportion of instruction is given to recitation exercises (Gall, 1970; Hoetker and Ahlbrand, 1969). At least two important elements of recitation techniques, questions and speaking before a gorup, have been found to be stressors which create anxiety (Kubis, 1948; Zajonc, 1966). However, all learners may not benefit equally from stress produced by recitation. Persons characterized as facilitators by the Achievement Anxiety Test (Alpert and Haber, 1960) retained more from reciting than debilitators, whose retention decreased as the rate of recitation increased (Schultz, 1970). The relatively poor performance of debilitators may be due to their tendency to "overload" short-term storage with intrusions of irrelevant thoughts associated with anxiety (Sarason, et al., 1960). In the case of recitation, these intrusions may include thoughts brought on by the learner's anticipation of being called upon to recite and by his relief after reciting or not being called upon.

The purpose of the present experiment is to identify instructional treatments which reduce the debilitating effects of intrusions created by recitation. One possible treatment is to minimize the threat to

[1] The authors acknowledge the assistance of the students and administration of the Altoona High School, Altoona, Pennsylvania. In particular, they are indebted to the gracious cooperation extended by Mrs. Grace Epright, Head Guidance Counselor.

self-esteem posed by recitation and thereby make the occurrence of intrusions less likely. Presumably, the threat to self-esteem is intensified when the learner feels his actions are being evaluated and reduced when evaluation is minimal. Accordingly, intrusions may be directly related to the extent to which the learner is "put on the spot" in the recitation situation. For the purposes of the present research, it was assumed that a sustaining instructional climate which was designed to support the learner is less of a threat to self-esteem than a directive climate which was designed to emphasize the evaluation of the learner by peers and an authority.

It also may be possible to override the disruptive effects of intrusions which do occur by requiring the learner to encode instructionally relevant information. Encoding, transforming input, or otherwise relating new responses to existing elements in the learner's cognitive structure is necessary to transfer information from short-term memory store to long-term storage (Atkinson and Shiffrin, 1968). One way of requiring encoding is to have the learner put the answer in his own words. As a consequence of translating, the learner must relate the new response to existing elements in his cognitive structure. Thus, the act of translation may require processing information so it is transferable to and retrievable from long-term storage. On the other hand, verbatim repetitions emphasize rehearsal rather than encoding and are, thereby, more subject to decay and the disruptive effects of intrusions.

Based on this rationale, a major expectation of this study is that translation in a sustaining climate facilitates retention of debilitators compared to debilitators in other conditions.

Method

Design

Each S in a group of six responded to three recitation questions by finding and reciting answers from printed textual material. Half of these Ss were selected as facilitators based on their extreme scores on the Achievement Anxiety Test (Alpert and Haber, 1960) while the remainder were judged as debilitators. Climate was varied by emphasizing or minimizing the evaluation of the Ss responses before the group and by withholding or offering praise after each response. These manipulations were designed to produce a directive or a sustaining climate. Half of the Ss in each climate condition were instructed to make verbatim responses while the remainder were told to translate the answers into their own words. These manipulations imply a 2 x 2 x 2 factorial analysis of variance design with two levels of climate (directive and sustaining), two levels of response mode (verbatim and translation), and two personality types (debilitators and facilitators).

Subjects

The Ss were 72 seniors from a local high school who were drawn from a larger pool of students who took the Anxiety Achievement Test (AAT). Two groups were defined by their extreme scores on the AAT as follows: debilitators were high scorers on the debilitating scale ($\overline{X} = 34.17$) and low scorers on the facilitating scale ($\overline{X} = 19.33$) while facilitators were low scorers on the debilitating scale ($\overline{X} = 20.79$) and high scorers on the facilitating scale ($\overline{X} = 29.42$). Three facilitators and debilitators were randomly assigned to the recitation sessions. When students who were randomly assigned to an experimental session were unavailable due to absence or scheduling conflicts, other students were assigned to take their place. The scores of these "fillers" were not included in the analyses. The missing students were reassigned to a later session.

Experimental Materials

The experimental materials were identical to those used in an earlier study (Schultz, 1970). Briefly, they consisted of a set of 36 slides each of which contained a sentence or brief paragraph which described an experiment on attitude change. Printed versions of the slides were used by _Ss_ to identify answers to 18 recitation questions asked by _E_. These questions were factual in nature and easily answered from the printed material.

Procedures

The experiment was described to _S_ as a lesson with three parts: a brief slide presentation, a recitation period in which _S_ would search for and recite answers to questions asked by _E_, and a test on the information presented during the experimental lesson. During the first stage, 36 slides were projected at eight second intervals providing only enough time for rapid scanning of their contents. The slides contained a modified description of the Festinger and Carlsmith (1959) study of the cognitive effects of forced compliance. The original version was modified to make it more understandable to high school students. Printed versions of the slides were then distributed for use during the recitation period. Before the recitation period, _Ss_ received instructions designed to induce the experimental conditions.

Treatments. The purpose of the directive climate was to emphasize the evaluation of the learner during recitation. These Ss were instructed as follows:

> It is important that you understand why you are reciting in this lesson. The reason we are asking you to recite is to find out whether you know or at least can find the right answer. You'll help us by making sure your answer is correct.

The sustaining treatment was designed to minimize the evaluation of the learner during recitation. These Ss were instructed as follows:

> It is important that you understand why you are reciting in this lesson. The reason we are asking you to recite is not to find out whether you know or can find the answer, but just to share your answer with the others and to make sure every one gets the same answers. You'll help us by sharing your three answers with the rest of the group.

Each correct answer given by Ss in the sustaining condition was verbally reinforced by E who responded with, "Good," "Right," "That's it," etc. When a correct answer was given by Ss in the directive condition, E merely went on to the next recitation question. In both conditions, when S made an error, he was asked to reexamine the printed material.

The climate conditions were orthogonally crossed with the two modes of response treatments. Half of the Ss in each climate condition were given verbatim instructions. They were told to repeat the answers in the same words as the printed material without making any changes. The rest were given translation instructions to the effect that they must rephrase the answers in their own words rather than to repeat them as they were in the printed material.

Measures. After the recitation period, the first ten items of the State Anxiety Test (Spielberger, Gorsuch, & Lushene, 1968) were administered. The test was represented as a "Self-Evaluation Questionnaire" in which *Ss* were to describe how they felt during the recitation period. Retention was measured with a 30 item multiple-choice test which provided indicies of intentional learning (information included in answers to recitation questions), incidental learning (information not recited but included on the slides and printed material), and recitation learning (answers to the questions *S* recited). Before giving the test, *E* attempted to reduce anxiety posed by the test so that any differences which were obtained could be attributed to the recitation period. They were given the following instructions:

> This is more a test of different teaching techniques than it is of you. Unlike the recitation period where everyone could see how well you did, your individual test score will not be known to others and is of nc interest to us. Your score will be only one of many that are averaged together and lost among other scores.

After the test, *Ss* completed items 11 through 20 of the State Anxiety Test, this time using the completion of the multiple-choice test as a reference point.

Results

The first ten items in the State Anxiety Test were used to determine whether the climate conditions were induced. The analysis of variance of scores derived from these items yielded $F\ (1,64) = 1.25$, $p > .05$ for the effect due to climate. Although this difference was not significant, the mean score of the sustaining condition ($\overline{X} = 18.75$) was lower than that of the directive condition ($\overline{X} = 20.03$). Thus, the direction of the means was suggestive of induction. A similar analysis of State Anxiety yielded $\underline{F} < 1.0$ for the effect due to response mode. The verbatim response mode ($\bar{X} = 19.36$) differed little from the translation condition ($\bar{X} = 19.41$). While deficiencies in response mode were not reflected in differences in anxiety, personality differences were. An analysis of variance yielded $\underline{F}\ (1,64) = 14.36$, $\underline{p} < .001$ in which facilitators ($\bar{X} = 17.31$) reported less anxiety than did debilitators ($\bar{X} = 21.47$).

The last ten items of the State Anxiety Test were administered immediately after the multiple-choice test. The purpose of this administration was to determine whether any differences in retention which were obtained between the experimental conditions were due to the effects of stress produced by recitation or to stress produced by the test itself. No differences were obtained between the sustaining ($\bar{X} = 19.25$) and directive climate ($\bar{X} = 19.64$) or between the verbatim ($\bar{X} = 19.75$) and translation ($\bar{X} = 19.14$) response modes. An analysis of variance of scores from these items yielded $\underline{F} < 1.0$ for the effect due to both climate and response mode treatments. An analysis of the effect due to personality yielded $\underline{F}\ (1,64) = 24.34$, $\underline{p} < .001$, in which facilitators ($\bar{X} = 16.80$) once again reported less tension than debilitators ($\bar{X} = 22.09$). These findings suggest that differences in retention due to climate may be the result of stress from the recitation period, but do not appear to be due to the stress associated with test-taking. On the other hand, it is not clear whether any main effects in retention obtained between facilitators and debilitators would be due to stress from recitation of from the test situation.

In addition to the overall retention score, three sub-scores of that measure were also analyzed. These included intentional learning, incidental learning, and the learning of answers to items S personally recited. An analysis of variance of overall retention score yielded F (1,64) = 2.92, p < .09 for the effect due to climate, F (1,64) = 2.30, p < .13 for the effect due to response mode, and F (1,64) = 2.50, p < .12 for the effect due to personality. For each analysis, the effects were above traditionally accepted levels of significance (p = .05) but in each case, the means were in the expected direction. The sustaining scores ($\bar{X}$ = 14.63) were higher than the directive scores ($\bar{X}$ = 13.17), translation ($\bar{X}$ = 14.56) higher than verbatim ($\bar{X}$ = 13.25), and facilitation ($\bar{X}$ = 14.58) higher than debilitators ($\bar{X}$ = 13.22).

Scores on items which tested intentional learning were analyzed for the effect of climate, response mode, and personality. The effects of both climate and personality were not significant (F [1,64] = 1.89, p > .10 for the effect due to personality and F < 1.0 for the effect due to climate). A similar analysis yielded F (1,64) = 3.88, p < .05 for the effect due to response mode. Retention scores of the translation group ($\bar{X}$ = 8.94) were higher than those of the verbatim group ($\bar{X}$ = 7.67). The multiple-choice test included 12 items which tapped incidental learning, i.e., they required information included on the printed material, but which Ss were not directed to examine by the recitation questions. An analysis of variance of incidental learning yielded F (1,64) = 5.18, p < .03 for the effect due to climate. Retention for the sustaining group ($\bar{X}$ = 6.09) was higher than for the directive group ($\bar{X}$ = 5.11). Neither the response mode nor personality effects were significant (F < 1.0).

A further analysis was made to determine how well S retained answers to questions they personally recited in the experimental "class." It will be recalled that each S responded to only three questions and that a different random assignment of questions was made for each experimental session. This analysis yielded $F\ (1,64) = 4.88$, $p < .03$ for the effect due to response mode, in which retention by the translation group ($\bar{X} = 2.03$) was greater than by those who made verbatim repetitions ($\bar{X} = 1.56$). A similar analysis yielded $F\ (1,64) = 2.85$, $p < .10$ for the effect due to personality. Facilitators ($\bar{X} = 1.97$) retained more than debilitators ($\bar{X} = 1.61$). The effect due to climate yielded $F < 1.0$.

An earlier study (Schultz, 1970) found that facilitators scored higher than debilitators on measures of intentional and incidental learning. The differences in overall retention in the present study were consistent in their direction, but marginal in magnitude. Since the directive-verbatim (DV) condition was closest to that of the earlier experiment, a pairwise comparison of overall learning was made between facilitators and debilitators in that group. This analysis yielded $t\ (64) = 1.16$, $p > .10$ for the effect due to personality ($\bar{X} = 13.56$ for facilitators and $\bar{X} = 11.56$ for debilitators). An analysis of incidental learning yielded $t\ (64) = 1.97$, $p < .05$ for the effect of personality in which facilitators ($\bar{X} = 5.44$) scored higher than debilitators ($\bar{X} = 4.67$).

The primary concern of the present experiment was with the retention of debilitators. It was hypothesized that both a sustaining climate and translation would improve their retention. In order to test this hypothesis, a separate analysis was made of the retention scores of debilitators in a 2 x 2 factorial analysis of variance design with two levels of climate and two levels of response mode. The results can be summarized as follows: (1) The analysis of overall learning scores yielded $\underline{F}$ (1,32) = 2.10, $\underline{p}$ < .16 for the effect due to climate and $\underline{F}$ (1,32) = 2.38, $\underline{p}$ < .13 for the effect due to response mode; (2) the analysis of intentional learning yielded $\underline{F}$ < 1.0 for the effect of climate and $\underline{F}$ (1,32) = 3.27, $\underline{p}$ < .08 for the effect due to response mode; and (3) the analysis of incidental learning yielded $\underline{F}$ (1,32) = 5.19, $\underline{p}$ < .03 for the effect due to climate and $\underline{F}$ < 1.0 for the effect due to response mode. The means used in this analysis are summarized in Figure 1 where it may be seen that both a sustaining climate and translation facilitate retention for debilitators. However, the beneficial effects of a sustaining climate on the retention of debilitators is felt primarily on incidental learning while the effect of response mode is primarily on intentional learning.

One implication of the present hypothesis is that the greatest differences in retention among debilitators would occur between the ST condition where both remedial treatments were experienced and the DV condition where neither treatment was experienced. The direction of the means, summarized in Figure 1, is consistent in this regard. Of the four cells containing debilitators, scores of the ST group are highest and those of the DV group the lowest on each measure. A pariwise comparison of the ST and DV debi litators yielded $\underline{t}$ (32) = 2.12, $\underline{p}$ < .05 for overall learning. As represented in Figure 1a, retention of debilitators in the ST condition is not only superior to that of other debilitators, particularly those in the DV condition,

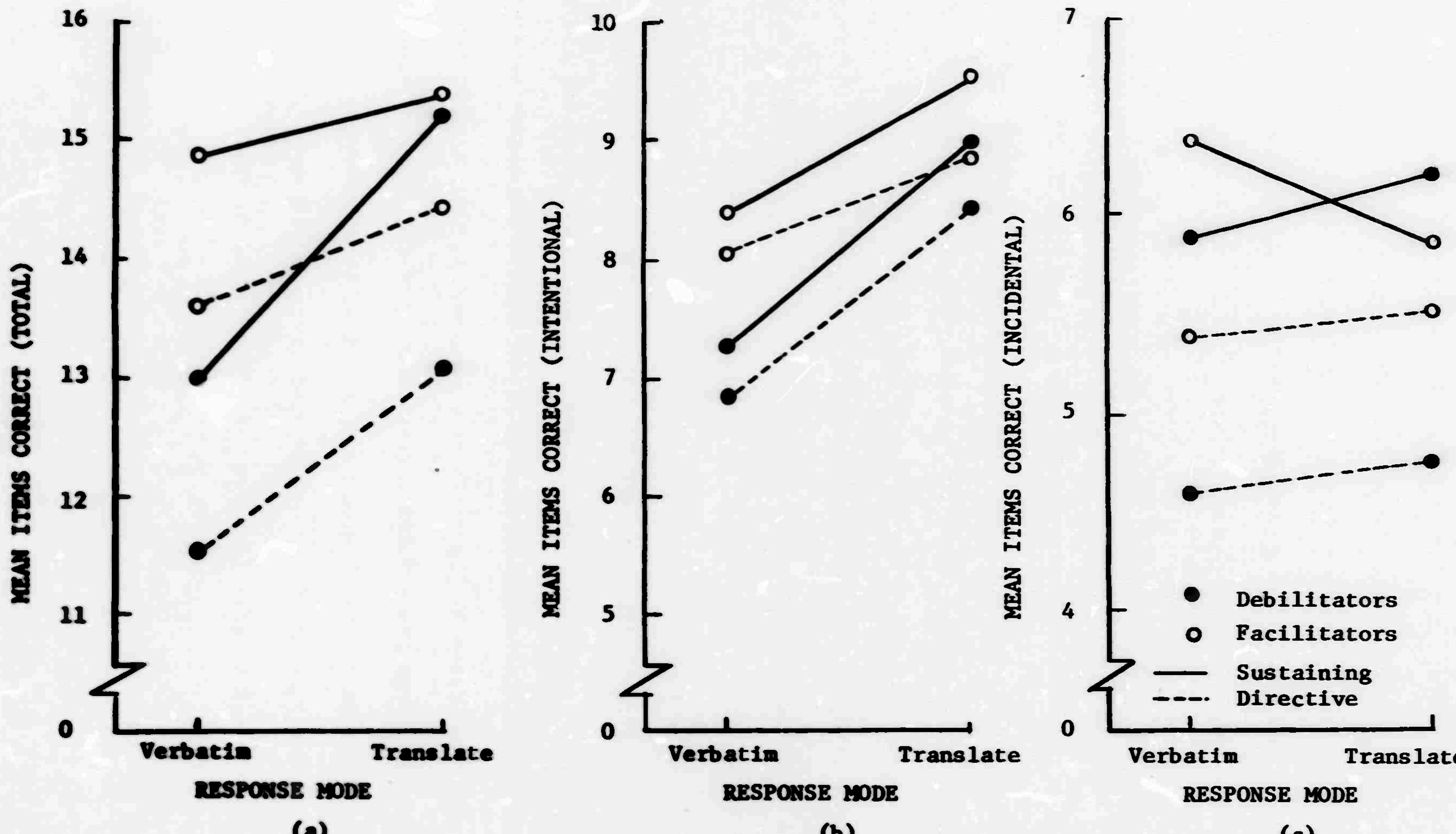

Figure 1. Performance of facilitators and debilitators in two response mode and climate conditions on measures of retention.

it approximates the retention of facilitators. The same trend was obtained in pairwise comparisons of retention sub-scores. These analyses yielded $\underline{t}$ (32) = 1.68, $\underline{p}$ < .05 (one-tailed) for intentional learning and $\underline{t}$ (32) = 1.97, $\underline{p}$ < .05 (one-tailed) for incidental learning.

Discussion

Both climate and response mode are variables which clearly have important effects on retention in recitation settings, particularly for persons classified as debilitators. Although the sustaining climate resulted in moderately higher overall retention scores than the directive climate, its most pronounced effect was to facilitate incidental learning. This finding is important for instructional purposes because teachers typically require or encourage learning beyond the target information to which they explicitly direct their learners.

The superior incidental learning of the sustaining group may have been due, in part, to the reduced ego threat they experienced. As a result they were less concerned with committing errors and, therefore, free to "stray" from the instructor's definition of the content. On the other hand, Ss in the directive condition may have felt compelled to restrict their learning to the recitation questions. In some regards, questions inserted before relevant textual material function as the written counterparts of recitation questions. Frase's (1970) review of the effects of pre-questions in textual material suggests that, in general, they depress incidental learning. However, the tendency of pre-questions to produce greater selectivity in learning was overcome under conditions of high motivation (financial payment was used as an incentive). Thus, another factor which may contribute to the higher incidental learning by the sustaining group in the present experiment is the motivational effect of E's verbal reinforcement which may have been comparable to the incentive of modest financial payment used in the studies Frase reviewed.

As might be expected, translation had little effect on incidental learning. However, translation appears to be an important variable influencing how much is retained of the information S personally recites (recitation learning) as well as influencing how much information recited in "class" by others (intentional learning) is retained. To maximize transfer from short-term to long-term storage, it appears to be important to go beyond rehearsal (verbatim responses) and to require the learner to relate new information to previously acquired information. Presumably, translation is one way of ensuring encoding and thereby, the transfer of information to long-term storage where it is maintained and retrievable.

The performance of debilitators in the ST condition was of particular interest. The findings support the hypothesis that translation in a sustaining climate improves retention of debilitators compared to other debilitators who received no remedial treatment. The direction of the means of debilitators in the four treatment conditions was consistent across all measures of retention. Those of the ST treatment were invariably the highest while those of the DV group, which received no remedial treatments, were the lowest. Although these findings do not explain the hypothesis that short-term storage is vulnerable to disruption from irrelevant thoughts, they are consistent with it. Presumably, a sustaining climate reduces the number of intrusions and encoding mitigates the effects of those intrusions which do occur. In this regard, encoding functions in the recitation setting as Sieber, Kameya, and Paulson (1970) found memory supports to function with a problem-solving task. In both cases, the disruption of short-term memory which debilitators appear to be particularly susceptible to is overcome.

The main effect of personality type (facilitators or debilitators) is less clear than the effect of either climate or response mode. The differences in retention of very moderate magnitude which were obtained may have been due to the debilitative effects of stress induced by test-taking as well as by reciting. The contribution of each to the differences in retention which were obtained is not clear from the present analyses. Moreover, the relatively large differences in retention between facilitators and debilitators obtained in an earlier study (Schultz, 1970) were not found in the present research. The failure to replicate earlier findings may be accounted for in part by the facilitative effect of translation and a sustaining climate on the retention of debilitators.

The findings of the present experiment, together with those of an earlier study (Schultz, 1970), permit the following conclusions regarding recitation strategies: (1) More of the instructional topic is retained when learners do not know when they will be called upon to recite than when their turn is known, (2) Moderate rates of reciting result in greater retention than frequent reciting, (3) Facilitators learn more from reciting than debilitators, (4) A sustaining climate facilitates retention, particular of incidental learning compared to a directive climate, (5) Translation facilitates retention, particular of intentional learning compared to verbatim responding, and (6) Translation in a sustaining climate improves intentional and incidental retention of debilitators compared to those who experience a directive climate and a verbatim response mode.

References

Alpert, F., & Haber, R. N. Anxiety in academic achievement situations. Journal of Abnormal and Social Psychology, 1960, 61, 207-215.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In Spence, K. W., and Spence, J. T. (Eds.), The psychology of learning and motivation: Advances in research and theory. Vol. 2. New York: Academic Press, 1968.

Festinger, L., & Carlsmith, J. M. Cognitive consequences of forced compliance. Journal of Abnormal and Social Psychology, 1959, 58, 203-210.

Frase, L. T. Boundary conditions for mathemagenic behaviors. Review of Educational Research, 1970, 40, 337-347.

Gall, M. D. The use of questions in teaching. Review of Educational Research, 1970, 40, 707-721.

Hoetker, J., & Ahlbrand, W. P. The persistence of recitation. American Educational Research Journal, 1969, 6, 145-167.

Kubis, I. F. Adaptation of the psychogalvanic response to a visual, auditory, and ideational stimulus. American Psychologist, 1948, 3, 256.

Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R. R., & Ruebush, B. K. Anxiety in elementary school children. New York: Wiley, 1960.

Schultz, C. B. Recitation strategies: The effects of rates and schedules of verbal responses on retention. In F. J. Di Vesta, D. R. Peters, N. M. Sanders, C. B. Schultz, and P. D. Weener. Instructional strategies: Multivariable studies of psychological processes related to instruction. A semi-annual report to the Advanced Research Projects Agency, Order Number 1269, January, 1970.

Sieber, J. E., Kameya, L. I., & Paulson, F. L. The effect of memory supports on the problem-solving ability of test-anxious children. Journal of Educational Psychology, 1970, 61, 159-168.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. The state-trait anxiety inventory. Tallahassee, Florida: Psychology Department,

Zajonc, R. B. Social psychology: An experimental approach. Belmont, California: Wadsworth, 1966.

The Effects of Subjective Organization Instructions and Verbal Creativity On the Recall of Random and Organized Lists

Paul Weener and Ovid Tzeng

When faced with the task of remembering many separate bits of verbal information, prople tend to group or cluster the bits into subjective units or clusters. The semantic, syntactic, and phonemic features which define words are of such a number that almost any set of "unrelated" words or grapheme units can be partitioned into subsets within which items share a number of features (Bower, 1970). Verbal learning studies have shown that the dominant strategy of subjects in free recall tasks is to create word clusters and to use these clusters as recall units (Bousfield, Puff, & Cowan, 1964; Tulving, 1962).

In instructional settings, students impose organization on verbally presented material regardless of whether or not that material has been preorganized by the instructor or a textbook. The student makes use of the organizational features of the material which have been taught to him and are shared in common by his cultural or scientific community. In addition, the student adds unique organizational features as well. In this sense, subjective organization can be thought of as consisting of two components, (a) the normative, or common component which is imposed by the textbook or instructor and (b) an idiosyncratic component which is based on a given individual's unique structuring response to the material. The "meaning," or structural relations, among the

PRECEDING PAGE BLANK

elements of information has a component which is common to all the people in a given social community because of the common experiences and uses which members of the same culture have with these elements. In addition to this common component is an idiosyncratic component which reflects the unique idiosyncratic experiences, contexts, and uses with which the elements to be learned have been associated.

It can be argued that students will learn best when they are free to impose their own organizational schemes on material to be learned. Mandler & Pearlstone (1966) found that subjects in a free organization group took half as many trials to obtain stable groupings of 52 "unrelated" words when compared to a constrained group which was required to use the groupings provided by subjects in the free organization group. Deagman (1969) showed a similar effect in a free recall task using three different kinds of word lists. Subjects who were free to group the words freely on a sorting board later recalled more words than those subjects who were constrained to use the categories which were used by the subjects in the free organization group. This was true for word lists which had high and moderately salient conceptual categories built into the lists and for a list of randomly selected words.

Subjective organization can be viewed as a mathemagenic behavior, involving all three aspects of mathemagenic behavior as described by Rothkopf -- translation, segmentation, and processing. Rothkopf's disillusionment with the "calculus of practice" approach to the study of instruction resulted in his statement of a "practice indeterminacy principle": "A particular, objectively described practice event can result in several different psychological states" (1968, p. 112). The research in subjective organization emphasizes these differences in psychological states and attempts to describe their effects.

The purpose of the present study is to investigate the effects on learning of imposing constraints on a subject's freedom to categorize words. It is hypothesized that the facilitative effects of free organization activities will be greater for unstructured word lists than for word lists which have "built-in" normative categoreis. The rationale for this hypothesis is that unstructured word lists provide greater opportunity for a person to employ subjective organizational strategies. Word lists which consist of highly common sub-categories of words should be organized similarly for most subjects, and the negative effects of being forced to use another person's categorization scheme should be less in such instances than in situations which produce highly idiosyncratic word groupings. The study was also designed to investigate the relationship between a measure of verbal creativity and recall under the different experimental conditions.

Method

Design

The study was carried out as 3 x 2 factorial design with organizational mode and the category saliency of the stimulus word lists defining the two factors. The three organizational modes were (a) free, (b) constrained, and (c) constrained random, and the word lists were classified as high and low saliency.

Subjects

Sixty-six subjects from the introductory educational psychology course participated in the experiment. They were awarded points toward their final course grade for participation in the experiment.

Materials

Two sets of 40 stimulus words were constructed. The list of high category saliency words were selected from the norms provided by Battig & Montague (1969). These norms were constructed by asking 442 subjects to give as many specific responses to a generic category label as they could in a 30-sec. period. Eight words were selected from each of five categories from responses which were given by at least 10 subjects. The five categories used were (a) a type of footgear, (b) a weapon, (c) a carpenter's tool, (d) a natural earth formation, and (e) a part of a building. The low category saliency list of 40 words was then constructed by selecting paired words from the Thorndike and Lorge (1959) general word count which matched the words in the high saliency list for frequency, number of syllables, and first letter. All words in both lists were nouns.

The Remote Associates Test (RAT) was developed to measure "the ability to think creatively" (Mednick & Mednick, 1967). The theory on which the test is based describes the creative thinking process "as one of seeing relationships among seemingly 'mutually remote' ideas and forming them into new associative combination" (p. 1). A subject is required to supply a response word which has some association with each of three stimulus words, e.g., "scotch" is the correct response to the three stimulus words, "soda," "kilt," and "butter." Alternate form reliability reported by Mednick & Mednick (p. 13) was .81 based on a sample of 71.

Procedure

Subjects were randomly assigned to one of the six experimental conditions and participated individually in the experiment. After being seated at a table, they were instructed to put on a set of headphones and that all instructions were taped and would be presented through the headphones. In front of the subject on the table was a 36" x 27" piece of cardboard ruled into nine equal rectangles. A pack of 40 plain 3" x 5" cards with one word per card were placed face down on the cardboard. The high and low category saliency groups differed only in the words which were printed on the cards.

The taped instructions for the free organization group were as follows:

> The experiment you are about to participate in investigates how people organize and recall verbal materials. Forty 3 x 5 cards are face down on the table in front of you. On each card is written one word. Your job will be to turn the cards over one at a time, pronounce the word out loud, and place the card in one of the spaces on the sorting board. Place the words which seem to go together in the same space on the sorting board. You can use as many of the spaces as you wish. You can use any grouping of the words which will help you remember the words, except - do not group the words alphabetically, but rather group the words based on the meaning of the words.
>
> You will turn the cards over one at a time and have 5-secs. to place the word on the sorting board. After you start a bell will ring every 5-secs. to pace your work. Every time the bell rings you must place a card on the sorting board and pick up a new card. After all the cards have been placed, you will be given another set of cards which contain the same 40 words in a different order. You will sort them once again following the same procedure. After the second sorting, you will be asked to write down as many of the words as you can. Remember, sort the words into meaningful categories which will help you to remember the words.

The constrained and constrained random group followed a similar procedure but in both of these conditions a number from 1 to 9 was written in the upper right hand corner of the card corresponding to numbers written in the 9 spaces on the sorting board. Each subject in the constrained group was yoked to a subject in the free organization group. The number on the card corresponded to the space in which that word had been sorted on the second trial by the yoked subject in the free organization group. In the constrained random group, the number of categories and the number of words per category were the same as the yoked subject but the words were assigned at random within each category. The instructions to these two groups were the same and the procedure was similar to the gree organization group. The significant part of the instructions which distinguished the two constrained groups from the free group was as follows:

> Your job will be to turn the cards over one at a time and pronounce the word out loud. Then look at the number written on each card and place the card on the sorting board in the category with the same number. Try to make some sense out of the groupings of words so that you can remember the words better.

After a subject had sorted the deck for the second time, he was asked to write down as many words as he could on a recall sheet. The 8 1/2" x 11" recall sheet was marked off into nine rectangles like the sorting board. All subjects were instructed: "Try to write the words into the same groupings that you used when you sorted them." They were given five minutes for recall. Following recall, the Remote Associates Test was administered following standard procedures with a 30 minute time limit.

Results

A 3 x 2 analysis of variance was carried out to determine the effects of organizational mode and type of word list. The means and standard deviations for number of words recalled in each of the six conditions is presented in Table 1.

The main effect due to organizational mode was significant, $\underline{F}$ (2,60) = 31.9, $\underline{p}$ < .001. The main effect due to word lists was not significant, $\underline{F}$ (1,60) = 3.45, $\underline{p}$ > .05, and the interaction between organizational mode and type of word list was not significant, $\underline{F}$ (2,60) = 1.28, $\underline{p}$ > .05. A Scheffe's test on the three means for the different organizational modes indicated that all three means were significantly different from each other ($\underline{p}$ < .05).

An aptitude-treatment interaction was hypothesized between verbal creativity and the organizational mode. In order for this hypothesis to be supported, the correlations between verbal creativity and number of words recalled would be different for the treatment conditions. The results of the correlation and regression analysis are presented in Table 2.

A test of the aptitude-treatment interaction hypothesis was carried out by testing the null hypothesis that the three regression slopes were equal (Cronbach & Snow, 1969). The null hypothesis was not rejected, $\underline{F}$ (2,60) = 1.74, $\underline{p}$ > .05. Although the differences in the regression slopes were not significant, the correlation between RAT performance and recall was significantly different than zero in the constrained-random group. The correlation between number of words recalled and RAT score was -.04 for the combined free organization and

Table 1

Means and Standard Deviations for Number of Words Recalled

Word List		Organizational Mode: Free	Constrained	Constrained Random
High Saliency	$\overline{X}$	27.00	23.82	16.45
	SD	3.10	4.91	3.14
Random Saliency	$\overline{X}$	24.09	21.18	16.82
	SD	2.88	4.53	3.52

Table 2

Means and Standard Deviations of RAT Scores and Correlation and

Regression Coefficients of RAT Scores with Number of Words Recalled

	Organizational Mode: Free	Constrained	Constrained Random
Remote Associates Test			
$\overline{X}$	14.32	13.46	14.18
SD	4.97	4.48	4.71
r	-.05	-.11	+.48*
b	-.03	-.11	+.33*

*$p < .05$

constrained group. These are the two groups which sorted the words into meaningful categories. When the correlation between RAT scores and recall performance for this combined group is compared to the same correlation for the random constrained group, a significant difference results, $z = 2.00$, $p < .05$.

Discussion

The results indicated that subjective organization of independent bits of information facilitated recall of that information. The facilitative effects were similar when the stimulus words contained salient categories and when they were randomly selected. It had been hypothesized that the effects of the different organizational modes would be greater for the randomly selected word lists than for the lists in which the categories were salient. This was based on the expectation that subjects in the high saliency condition would utilize the provided category groupings with a resulting small idiosyncratic component in the resulting word groupings, and that subjects in the random word list condition would have a large idiosy ncratic component in their word groupings. The larger the idiosyncratic component in the word groupings formed by subjects in the free organization condition, the less "helpful" the imposed word groupings would be for the yoked subjects in the constrained condition. What seemed to happen, however, is that the categories which were created by the free organization subjects from the random word list condition were as "helpful" to the yoked subjects as the categories created in the high saliency word list. The helpfulness of the imposed organization can be attested to by comparing the constrained to the constrained random group. Using someone else's grouping of the words resulted in considerably greater (32% more) recall than using the random groupings provided. The constrained and constrained random group followed identical experimental procedures so that the difference in recall can be attributed to the characteristics of the word groupings that they were constrained to work with in the sorting task.

The lack of a significant effect due to category saliency in the word lists was somewhat surprising. The variance accounted for by organizational mode was much greater than the variance accounted for by the categories which were or were not available in the word lists. Even though the words were selected randomly for the random word lists, subjects found meaningful categories of words imbedded in the list. These categories which subjects imposed on the random lists had sufficient inter-subject meaning to produce the large observed effect between the constrained and the constrained random group.

The differences among the three groups in the correlation between RAT score and number of words recalled seems to indicate a differential treatment effect. The RAT is designed to measure the ability to see "relationships among seemingly 'mutually remote' ideas and forming them into new associative combinations" (Mednick & Mednick, 1967, p. 1). The task for the constrained random group requires just such an ability. A person who scores high on the RAT should be able to perform better on a task in which word groupings are random, because such a person can form "new associative combinations" which will facilitate recall. In both the free organization and constrained organization group, the subject sorted the words into meaningful categories and thus the opportunity to utilize the abilities measured by the RAT are less available.

One possible implication of this finding is that some students require more pre-structuring of instructional materials than other students require. The RAT may be an appropriate measure of the relevant individual difference when the instructional materials are verbal.

The findings however do not support a differential treatment procedure. The regression lines relating RAT scores and recall do not cross; the free organization treatment is better than the constrained treatment which is better than the constrained random treatment in a free recall task for all levels of RAT. The differences in correlations implies that in an unstructured random verbal task, verbal creativity is correlated with recall performance whereas in a situation where the verbal materials are more organized, verbal creativity is not correlated with recall performance. Unstructured tasks, such as that presented to the random-constrained gorup, tended to "penalize" subjects with low RAT scores more than it "penalized" subjects with high RAT scores.

References

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories. Journal of Experimental Psychology Monograph, 1969, 80, No. 3, Pt. 2.

Bousfield, W. A., Puff, C. R., & Cowan, T. M. The development of constancies in sequential organization during repeated free recall. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 489-495.

Bower, G. Organizational factors in memory. Cognitive Psychology, 1970, 1, 18-46.

Cronbach, L. J., & Snow, R. E. Individual differences in learning ability as a function of instructional variables. Final report to U. S. Office of Education. School of Education, Stanford University, 1969.

Deagman, G. K. The effects of organizational mode and category saliency on the recall of verbal materials. Unpublished master's thesis, The Pennsylvania State University, 1969.

Mandler, G., & Pearlstone, Z. Free and constrained concept learning and subsequent recall. Journal of Verbal Learning and Verbal Behavior, 1965, 5, 126-131.

Mednick, S. A., & Mednick, M. T. Remote Associates Test. Boston: Houghton Mifflin, 1967.

Rothkopf, E. Two scientific approaches to the management of instruction. In Learning Research and School Subjects, Gagne, R., & Gephart, W. J. (Eds.). Itasca, Illinois; Peacock, 1968.

Thorndike, E. L., & Lorge, I. Teacher's word book of 30,000 words. New York: Teachers College, Columbia University, 1959.

Tulving, E. Subjective organization in free recall of "unrelated" words. Psychological Review, 1962, 69, 344-354.

Section III

Goal Expectancy as a Variable

in Teaching-Learning Situations

Summary

Type-of-Test Expectancy Effects on Learning of Word Lists and Prose Passages

Nicholas M. Sanders and Ovid Tzeng

Technical Problem

Students say they prepare for an essay type of test by studying interrelationships, while they prepare for objective tests by studying for specifics without regard to interrelationships. Empirical research has not convincingly demonstrated differences between the essay and/or objective test scores of learners preparing for essay tests and learners preparing for objective tests. The two studies reported here provided for more rigorous investigation of the effects of type-of-test expectancy with word lists and logically structured prose paragraphs.

General Methodology

In both investigations subjects studied and then were tested on three sets of materials. The materials in the first study were three 40-word lists, and in the second study the materials were three, seven-sentence paragraphs. Subjects were led to expect a test on either specific words (or sentences) or interrelationships among the specifics - serial order of words in first study and logical implications of sentences in the second study - by being appropriately instructed before studying each set of materials and by taking the respective type

PRECEDING PAGE BLANK

of test on the first two sets of materials. The major criteria were performances on tests for both types of learning (specifics and interrelationships) after studying the third set of materials.

Technical Results

There were strong type-of-test expectancy effects in the first study, with those persons expecting to recall words in any sequence actually recalling more words, but fewer words in the presented sequence, than those who expected to be required to recall words in serial order. In the second study there were no overall type-of-test expectancy effects. Instead, each of the three prose passages used resulted in a different relationship between performances on the two types of tests as a function of which type of test was expected.

Implications for Education and for Further Research

The difference in findings in the two studies may have been due to the extent to which appropriate preparation for the two types of tests conflicted with one another. If this explanation is appropriate, then these studies would imply that for most educational objectives one would anticipate no type-of-test expectancy effects.

However, another interpretation of the different findings in the two studies is that type-of-test expectancy effects are manifest with recall measures (as in the first study) but not with recognition measures (as in the second study). If this second interpretation were correct, there would be educational implications for recall measures, though not for recognition measures.

Thus, further research should be carried out to resolve the issue, most likely by using recall measures (as in the first study) with prose passages (as in the second study).

Type-of-Test Expectancy Effects on Learning of

Word Lists and Prose Passages

Nicholas M. Sanders and Ovid Tzeng

Students have reported that they study differently for different types of examinations (Douglass & Talmadge, 1934; Meyer, 1935; Silvey, 1951; and Terry, 1933). In general, they say they attempt to master specific details for objective tests and more comprehensive interrelationships for essay tests. However, studies in which the type of test anticipated has been experimentally manipulated have often not revealed any differences in learning outcomes (Hakstian, 1971; Vallance, 1947; and Weener, 1971). While Meyer (1934, 1936) was able to demonstrate learning outcome differences, the subjects in his studies were informed that they should study for a particular type of test and should not study for other types. Thus, a strong explicit demand for differences was possibly instituted in his studies. One other study (Sax & Collet, 1968) revealed that subjects expecting a "recall" test on applications did more poorly on a multiple-choice test of applications than did those expecting a multiple-choice applications test. However, Sax and Collet's "recall" condition subjects were older than their multiple-choice subjects, and age may be an index of several variables of importance in test taking behavior (Hakstian, 1971).

Though the conclusion that would seem to be implied from the available research is that there are no type-of-test expectancy effects on learning outcomes, the criteria used in the studies have never allowed for a comparison of outcomes in accordance with the student's claims concerning their study behavior. That is, the criterion tests have not been designed to compare knowledge of specifics and knowledge of interrelationships. If essay test scores are partially determined by the presence of specific points (as is often doen to increase the inter-rater reliability) or if some of the objective test items measure interrelationships among specifics by including distractors pertaining to other specifics in the material covered, the comparison of the scores on the tests of different format may reveal no differences. The purpose of the present studies is to determine whether expectation of a test on specifics will produce different learning outcomes from an expectancy of a test on interrelationships without regard to the usual distinctions of test format.

The first study represents a well-defined and controlled attempt to demonstrate type-of-test expectancy effects. Students studied word lists and were led to expect either a test on knowledge of specific words or a test on knowledge of interrelationships among the words. The conceptual basis for predicting test expectancy effects stem from some research in human memory. Tulving (e.g., 1964) and others have demonstrated that subjects asked to recall as many words as possible from a list of words engage in a process of regrouping the words. This process is called subjective organization, to indicate that the learner groups the words in a way most meaningful to himself. If memory for as many details as possible were the desired outcome of studying, subjective organization would be appropriate. However, if the desired outcome were memory of the words in the organization as presented, subjective organization might be antithetical to that goal. Type-of-test expectancy effects should therfore be manifest with these contrasting relevant study procedures, if the learner is capable of modifying his or her study procedures on the basis of the knowledge of which type of outcome is desired.

Experiment I

Method

Subjects. Thirty students enrolled in an introductory educational psychology course volunteered to serve as subjects and received extra grade credits in the course for their participation in the experiment. The subjects were administered the experiment in individual sessions, and were assigned to treatment conditions in an alternating fashion in the order in which they appeared for the session.

Materials. Three non-overlapping word lists were used in the study. Each list contained 40 words that were selected from eight categories, randomly chosen from the norms of Battig & Montagne (1969). Five words were chosen from the sixth to the eleventh response frequency ranks of each category. The words in each list were presented typed in a single vertical order down the center of a page of white typing paper, and were sequenced in a random order, with the restriction that no word followed another word from the same category.

A crossword puzzle, which was used as an interference task, involved 24 target words, each of which was drawn from the twelfth through twentieth response frequency ranking of each of the 24 categories used in the word lists. The cue phrase for target word was the corresponding category name. The puzzle was presented on a sheet of typing paper, with the cue phrases on the lower half of the page.

Procedure. All subjects were given four minutes to study each of the three word lists and were tested for recall immediately after the study of each list. The order of presentation of lists was counter-balanced so that one-sixth of the subjects had each one of the possible six sequences of the three lists.

Approximately one-half of the subjects in each list order condition were assigned to a free recall expectancy (FRE) treatment. These subjects were instructed prior to the study of first and second word lists that they were to remember as many words as possible. The instructions for the first two recall lists indicated that recall could be in any order in which they recalled the words, and the response sheet provided for the test was divided into a six by eight set of rectangles, which was designed to indicate to the subject that order of recall was unimportant. The subjects were allowed five minutes for the recall test.

The other subjects were assigned to a serial recall expectancy (SRE) treatment. These subjects were informed that they were to recall as many words as possible in the order in which they appeared on the stimulus list. The tests for their first two lists were preceded by instructions to recall the words in the presented order, and the response sheets for this group had two vertical columns of lined blanks with clear indications of which blank was to be used for recording the first word and how to continue recording words in the order the subjects recalled the words as having been presented. These subjects were also allowed five minutes for the recall test.

The instructions for study of the subject's third list were the same as they had been for the previous two lists, but upon completion of the study period all subjects were given both types of recall sheets and were instructed to recall both as many words in any order and as many words in serial order as they could. Recall time allowed for the third list was six minutes.

All subjects were then given ten minutes to work on the crossword puzzle described above in the materials paragraph. This task was designed to interfere selectively with the retention of subjects who had utilized category organization of the words to facilitate memory. To determine whether the anticipated interference occurred, subjects were retested for both free recall and serial recall of their third word list. Six minutes were again allowed for recall.

Each free recall response sheet was scored for the total number of words recalled. The serial recall sheets were scored by allowing one point for each word correctly following a preceding word; also, one point was allowed if a subject recorded the first word in the list on the designated line of the response sheet.

Results

Practice effects. To determine whether subjects recalled more on later lists than on earlier ones, a single-factor, repeated measures analysis of variance design was used. An F (2,28) = 6.82, $p < .01$ was computed for the practice effect on free recall (with FRE subjects only), while F (2,28) = 4.90, $p < .05$ was found for the analysis of serial recall data with SRE subjects only. The corresponding means and standard deviations are presented in Table 1, in which the constant

Table 1

Means and Standard Deviations of Two Types of Recall Scores

On First, Second and Third Word Lists[a]

	Word Lists		
Recall Type	First	Second	Third
Free Recall	26.7 (5.82)	28.5 (5.43)	31.0 (4.13)
Serial Recall	11.7 (6.06)	14.6 (5.21)	16.5 (6.85)

[a] Standard deviations are in parentheses.

improvement of recall from first list to last list is demonstrated for both free recall and serial recall. These effects may be due to warm-up and/or to the development of an expectation of, and consequent preparation for the particular type of recall test.

Treatment effects. Table 2 presents the means and standard deviations of free recall performances as a function of type of test expected and the within subjects variable, time of testing. The same statistics are given in Table 3 for the serial recall criterion. In addition to the two levels of test expectancy and the two testing times, the three word lists were included in a 2 x 3 x 2 mixed analysis of variance. Two such analyses were made, one for the free recall data and one for the serial recall data.

In both analyses type-of-test expectancy was a significant factor with F (1,24) = 59.12, $p < .001$ for the free recall measure, and F (1,24) = 92.98, $p < .001$ for the serial recall measure. Both differences were in the expected direction, with FRE producing superior free recall performances are SRE resulting in superior serial recall scores.

The time of testing, a repeated measurement, resulted in an F (1,24) = 6.90, $p < .05$, for free recall, and an F (1,24) = 4.12, $p < .10$, for serial recall. However, the interaction between type-of-test expected and time of testing was significant using free recall scores (F [1,24] = 8.12, $p < .01$), but not significant with the serial recall data (F [1,24] < 1.00). The nature of the interaction with the free recall measure was as expected: a greater loss in memory occurred in FRE than in SRE. The nature of the interaction using free recall scores and the lack of an interaction using serial recall scores indicates that

Table 2

Free Recall Means and Standard Deviations

As a Function of Type of Test Expected[a]

Type of Test Expected	Time of Testing on Third List: Immediate	After Ten Minutes Work on the Crossword Puzzle
Free Recall	30.6 (4.33)	28.6 (4.96)
Serial Recall	21.6 (5.86)	21.6 (5.82)

[a] Standard deviations are in parentheses.

Table 3

Serial Recall Means and Standard Deviations

As a Function of Type of Test Expected[a]

Type of Test Expected	Time of Testing on Third List: Immediate	After Ten Minutes Work on the Crossword Puzzle
Free Recall	2.9 (1.94)	2.3 (1.91)
Serial Recall	17.5 (6.98)	16.2 (6.64)

[a] Standard deviations are in parentheses.

the particular crossword puzzle task used interfered with subjective organization but not serial organization, and thereby demonstrates a source of interference when details are subjectively organized in memory.

Word lists were studied to determine the generalizability of findings across different materials. For the free recall data, the list main effect ($\underline{F}$ [2,24] = 2.19), the list X test expectancy interaction ($\underline{F}$ [2,24] = 2.00), and the list interactions involving time of testing (both $\underline{F}$'s [2,24] < 1.00) were not significant. List effects were revealed with the serial recall measure ($\underline{F}$ [2,24] = 3.72, $\underline{p}$ < .05), and the interaction between lists and type-of-test expectancy was significant ($\underline{F}$ [2,24] = 6.66, $\underline{p}$ < .01). However, the nature of these effects on the serial recall measures were not such as to qualify the type-of-test expectancy findings; the serial recall means for the three lists in the FRE being 1.8, 2.1, and 4.1, while the respective means in the SRE were 14.3, 21.6, and 11.4. There were no list interactions involving time of testing with the serial recall data; both relevant $\underline{F}$'s were less than 1.00.

Experiment II

Experiment I demonstrated that expectation of, and consequent preparation for, a particular type of test is a potent variable. In contrast to many of the previous research studies, analysis of the two types of tests involved allowed for a definitive conceptualization of the conflicting requirements of the two types of tests; one type allowed for reordering of words in recall (subjective organization), whereas the other type (serial recall) actually prevented any reordering. The subjects initially seemed to understand this, and with practice became more proficient at producing the outcome appropriate to each type of test.

The second experiment was an attempt to extend the findings of the previous experiment to materials more similar to school materials; instead of word lists, prose paragraphs were used. In this second experiment, the specifics that subjects were led to expect on the test were particular assertions made explicitly in sentences in the paragraphs. The interrelationship type of test was constructed from logical implications of sentences in the paragraphs and never were these questions answerable by reference to any single sentence presented.

In this experiment the two types of tests would seem to be differentiable as two levels of a hierarchy, with the knowledge of interrelationships among specifics being based on knowledge of the specifics. Of course, knowledge of specifics would not require knowledge of interrelationships among the specifics. However, it is possible that the learner who expects an interrelation type of test might forget the component specifics after mastering the interrelationships among them. And it is also possible that the learner expecting a specific test might be able to deduce the interrelationship between specifics as he or she confronts the unexpected question on the test. While this second experiment does not allow for a conceptual presentation of a clear conflict between different test expectancies, it is the type of experiment required for further definition of the type-of-test expectancy effect in the educational setting.

Method

Subjects. Sixty students from an introductory educational psychology course volunteered to serve as subjects and received extra grade credits in the course for their participation in the experiment. The subjects were randomly assigned to treatment conditions.

Materials. Three, seven-sentence paragraphs were selected as fulfilling three requirements. First, the content of the paragraphs was fictional, and therefore the likelihood of different knowledge backgrounds of subjects affecting their performance was decreased. Second, a rough comparison of the present study and other similar research was enabled by using two paragraphs Frase (1969, 1970) has studied. Finally, four of the seven sentences in each paragraph enabled logical deductions beyond the specific information explicitly presented. A brief specification of the structure of the critical four sentences can be described from an example, one of the paragraphs used in the present study, entitled "Survey on the Use of Reolam:"

> Reolam is a new, commercially marketed product. A recent survey has indicated that people who buy Reolam live on the north side of town. It is a well-known fact that the north side of town is a high pollution area. The City Council has been studying the characteristics of areas that produce considerable pollution. They have found that people in these areas are very cleanly. The research includes many new findings. One is that cleanly people prefer light colors instead of dark ones for furniture.

The paragraph includes four assertions about five classes of people: 1. Those who buy Reolam (A) are those who live on the north side of town (B). 2. Those who live on the north side of town (B) are those who live in a high polluation area (C). 3. Those who live in a high pollution area (C) are those who are very cleanly (D). 4. Those who are very cleanly (D) are those who prefer light colors instead of dark ones for furniture (E). These four assertions about the classes may be logically expressed as: A is contained in B, B is contained in C, C is contained in D, and D is contained in E. The remaining three sentences do not assert any type of logical relationship, and are assumed to serve as fillers.

The two other paragraphs used have the same structure, and are entitled "Astronomical Discoveries" (Frase, 1970) and "The Fundalas of Central Ugala" (Frase, 1969). The Fundalas paragraph was rearranged for the present study so that, as with the other two paragraphs, the sequence of relationships presented was from least inclusive to most inclusive.

Two sets of eight true-false test items were constructed for each paragraph. One set, referred to as <u>reproductive</u> by Frase (1970), consisted of items concerning relationships directly expressed in the paragraph. Four of these were true (e.g., the sentences assering A contained in B and C contained in D), and four were false (e.g., the sentences asserting B contained in A and D contained in C). The other set included items based on deductive reasoning, called <u>productive</u> by Frase (1970). Of the six possible valid inferences, four were randomly chosen for each paragraph (e.g., the sentences asserting A contained in C and B contained in E). The four corresponding invalid inferences (e.g., the sentences asserting C contained in A and E contained in B) were used as false items.

<u>Procedure</u>. The materials along with instructions were presented in typed form in a loose-leaf binder. Each binder contained the three paragraphs with the test items immediately following the corresponding paragraph. Each test item was presented on a separate page, to prevent

subject's comparison of questions and his or her previous answers. The sequence of items was randomly determined for each paragraph. All answers were recorded on a single response sheet not included in the binder.

The subjects were informed that they were to consider each page for as long as they chose, but not to look back to previous pages or to look forward until they were ready to proceed to the next page. One-sixth of the subjects read the three paragraphs in each of the possible six orders.

One-half of the subjects in each paragraph order condition read, prior to each paragraph, a study hint which suggested that the nature of test items would be specific (reproductive) and were given examples of a valid and an invalid item, based on two sentences unrelated to any of the paragraphs. The subjects in this condition, <u>reproductive test</u> expectancy (RTE), also saw only reproductive type items on the first two paragraph tests.

The other subjects read study hints prior to each of the three paragraphs that indicated, in a similar manner as described above, that test items would be of the productive type. Only productive items were included in their tests on the first two paragraphs. This condition is the <u>productive test expectancy</u> (PTE) condition.

After reading their third paragraph, all subjects were given both productive and reproductive test items, arranged in random order. These two types of items were scored separately for analysis. For comparability with the Frase (1969; 1970) studies, the scoring procedure entailed subtracting the proportion of invalid items called valid (false alarms) from the proportion of valid items called valid (hits).

The experiment was administered in a large room, with provisions made for up to six subjects to work independently. Since the epxeriment was subject-paced, a difficult crossword puzzle was included after the third paragraph items to allow all subjects to complete the experiment before any subject left the room. Completion time for the experiment itself varied from about 10 minutes to about 35 minutes.

Results

Practice effects. In contrast to the previous experiment, subjects performed no better on later tests than on earlier ones. The related F (2,58) for the reproductive items (using RTE subjects only) was 1.34 (p > .10), while for the productive items F (2,58) < 1.00. Thus, there is no evidence for the gradual development of type-of-test expectations.

Treatment effects. A 2 x 3 x 2 mixed analysis of variance was computed for comparing the two levels of test expectancy, the three paragraphs, and two levels of the within subjects variable, reproductive versus productive test items. The results of this analysis are presented in Table 4.

In contrast to the previous study, there was no interaction of type-of-test expected with type of test items. The significant triple interaction indicates that the paragraphs manifest distinctly different relationships between type of test expected and type of test item. The relevant means for the triple interaction are presented in Table 5. The paragraph on "Astronomical Discoveries" revealed no relationship between the type of test expected and the type of test item, while the "Reolam" paragraph yielded results contrary to those expected. Only the passage on "The Fundalas" produced the expected direction of interaction,

Table 4

Summary of the Analysis of Variance of Reproductive and Productive Test Item Performances

Source	df	MS	F
Between Subjects	59		
Type-of-test expectancy (A)	1	.0021	< 1
Paragraphs (B)	2	.8187	6.08**
A x B	2	.0152	< 1
Error	54	.1346	
Within Subjects	60		
Type of test (C)	1	.2521	4.46*
A x C	1	.0187	< 1
B x C	2	.2239	3.96*
A x B x C	2	.2063	3.65*
Error	54	.0565	

* $p < .05$

** $p < .01$

Table 5

Mean Proportion Score for Reproductive and Productive Test Items as a Function of Type-of-Test Expected and Paragraphs Studied[a]

Treatment Variables		Criterion Variables	
Paragraph Studied	Type-of-Test Expected	Reproductive Test Items	Productive Test Items
"Astronomical Discoveries"	Reproductive	.22	.00
	Productive	.20	.00
"Reolam"	Reproductive	.30	.28
	Productive	.42	.18
"Fundalas"	Reproductive	.45	.35
	Productive	.25	.50

[a] Scores were assigned by subtracting the proportion of invalid items called valid from the proportion of valid items called valid.

with more reproductive items correct with RTE than PTE and more productive items correct for PTE subjects than for RTE subjects.

Though the type of test item produced a significant effect, that effect is also contingent upon the paragraph used. From Table 5 one may discern that the reproductive items were easier than the productive ones for all paragraphs except the one on the "Fundalas", which yielded a productive item mean higher than the reproductive item mean.

Discussion

Type-of-test expectancy produced no significant effects in Experiment II. Neither was there any evidence that subjects became more proficient with either reproductive or productive test items as a function of prior experience. The potent variable in this experiment was the materials variable. Not only did the paragraphs differ in overall difficulty, they also produced markedly different patterns of relationships between the type of test expected and the type of test given. Only with one paragraph was there a tendency for better performance on the type of test expected, and subjects studying one of the two other paragraphs tended to perform worse on the type of test expected.

Unfortunately, there are no readily available dimensions along which to compare the paragraphs in ways that might clarify their effects in the second experiment. However, it is important to note that the studies by Frase (1969; 1970) and the first study presented in this report did produce strong effects over and above any effects the materials may have had. Why were the test expectancy effects so strong in the previous study and so weak and affected by the materials variable in the present study?

One of two factors possibly leading to different test expectancy effects in the first and second experiments reported here is the extent to which preparation for one type of test conflicted with being prepared for the other type test. In the first study there was a direct conflict: Remembering the serial order of the words prevented the subjective reordering of words into categories that would have allowed for remembering more words, while the subjective reorganization according to category clusters certainly prevent memory of the serial order of the words. In the second study there was no clear conceptual basis for conflict of expectancies on performance: Studying for the productive type test items would require detailed consideration of the specific sentences, while the learner who had remembered the specific sentences could deduce answers to the productive type items as those items were encountered on the test. If conflict between the type of preparation needed for the types of tests - as described above - is a prerequisite for type-of-test expectancy effects, then the occasions in educational settings where one would postulate such dramatic conflict would be very few if any. Based upon this analysis, the type-of-test expected should not strongly affect learning in the overwhelming majority of educational settings.

However, a procedural distinction between the studies could have resulted in the differences in findings: the type of memory measure used. In the first study, subjects were required to <u>recall</u> words, while in the second study subjects were asked to <u>recognize</u> valid statements. Usually recognition measures are considered to indicate a more surface level of memory than recall measures. Thus, the difference in results might be interpreted to mean that type-of-test expectancies affect only the deeper levels of memory and processing. If this were the correct interpretation of the differences in findings, then the expectancy of tests of specifics or expectancy of a test of interrelationships would be an extremely important variable for educators to consider.

The resolution of these two basically different interpretations of the different findings in the two studies reported here requires additional empirical research. While the first interpretation is more consonant with the issues as they were conceptualized in this report, the consideration of the type of memory measure used has been very important in research on learning and memory. Studies employing recall measures as criteria for type-of-test expectancy effects with prose passage materials should provide an answer to the issue raised by the present studies.

References

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph Supplement, 1969, 80(3, Pt. 2).

Douglass, H. R., & Tallmadge, M. How university students prepare for new types of examination. School and Society, 1934, 39, 318-320.

Frase, L. T. Influence of sentence order and amount of higher level text processing upon reproductive and productive memory. American Educational Research Journal, 1970, 7, 307-319.

Frase, L. T. Structural analysis of the knowledge that results from thinking about text. Journal of Educational Psychology Monograph Supplement, 1969, 60 (6, Pt. 2).

Hakstian, A. R. The effects of type of examination anticipated on test preparation and performance. Journal of Educational Research, 1971, 64, 319-324.

Meyer, G. An experimental study of the old and new types of examination: I. The effect of the examination set on memory. Journal of Educational Psychology, 1934, 25, 641-661.

Meyer, G. An experimental study of the old and new types of examination: II. Methods of study. Journal of Educational Psychology, 1935, 26, 30-40.

Meyer, G. The effect on recall and recognition of the examination set in classroom situations. Journal of Educational Psychology, 1936, 27, 81-99.

Sax, G., & Collet, L. S. An empirical comparison of the effects of recall and multiple-choice tests on student achievement. Journal of Educational Measurement, 1968, 5, 169-173.

Silvey, H. M. Student reaction to the objective and essay test. School and Society, 1951, 73, 377-378.

Terry, P. W. How students review for objective and essay tests. The Elementary School Journal, 1933, 33, 592-603.

Tulving, E. Intratrial and intertrial retention: Notes toward a theory of free recall verbal learning. Psychological Review, 1964, 71, 219-237.

Vallance, T. R. A comparison of essay and objective examinations as learning experiences. Journal of Educational Research, 1947, 41, 279-288.

Weener, P. D. The effects of recall mode and recall interval expectancies on note-taking and recall. In Di Vesta, F. J., Sanders, N. M., Schultz, C. B., & Weener, P. D., Instructional strategies: Multivariate studies of psychological processes related to instruction. Semi-annual report to Advanced Research Projects Agency (Order No. 1269). January, 1971. Pp. 61-74.

Summary

The Effects of Recall Interval Expectancy and Note Taking On Immediate and Delayed Retention

Paul Weener and Sam Rock

Technical Problem

This study investigated the effects of different expected recall intervals under note taking and no note taking conditions on the immediate and delayed recall of instructional material. It was hypothesized that different test interval expectancies influence note taking activities and the recallability of instructional materials. The general rationale for the research is that students modify their overt and covert study activities depending on the length of the expected interval between study and the criterion test, and that the manipulation of these expectancies will influence the learning outcomes.

General Methodology

Eight different experimental conditions were created by completely crossing three two-level factors: (1) note taking - no note taking; (2) immediate test expectancy - delayed test expectancy; (3) immediate test - delayed test. Groups of four to eight subjects participated in the experiment. Each of the subjects was presented with instructions which stated that he was to read and study a short article, that he could or could not take notes, and that he would be tested immediately after studying or one week later. The material which was studied for

thirty minutes was a rather difficult passage dealing with principles governing the development of species.

Half of the subjects were tested both immediately and one week later; half were tested one week later only. The immediate test consisted of a free recall test and an essay test; the delayed test consisted of a free recall test, an essay test, and a multiple-choice test.

Technical Results

Scores were obtained on a free recall test, an essay test, and a multiple-chioce test for the eight experimental groups. In addition, the number of words of notes taken in the note taking groups was counted. No significant main effects for the test interval expectancy factor or the note taking factor were found on any of the three criterion tests. An interaction between expected test interval and actual test interval on the free recall test results supported the contention that students do modify their study behaviors depending on their expectancy of when the criterion test will be taken. In support of this contention subjects took about twice as many notes when they anticipated a delayed test as when they anticipated an immediate test even though they knew that they would not be able to use their notes at a later time.

The interaction between the test interval expectancy factor and the note taking factor was significant for the multiple-choice test results, and approached significance in the analysis of the free recall results. The implication of these findings is that note taking seems to interfere with the learning strategies employed by a person preparing for an immediate test, but seems to facilitate the learning strategies employed by a person preparing for a test in the future.

Educational Implications

The results imply that note taking has little value in situations in which written instructional materials are presented and the notes taken cannot be used in later review. The results further imply that the recallability of material at a later time is dependent on the expectancy that a learner had when he is studying the material. If recall of material is required at a time in the future the learner should be so informed in order that the necessary adaptive study behaviors can be employed.

Implications for Further Research

Replication is needed to determine the reliability of the effects observed. The research should also be expanded to include comparisons of situations in which review of notes is permitted and to situations in which material is auditorily presented.

The Effects of Recall Interval Expectancy and Note Taking
On Immediate and Delayed Retention

Paul Weener and Sam Rock

The manner in which a student actively operates on visual or auditory stimulus materials in an instructional setting is dependent on his expectancies of when and in what form the information will have to be retrieved. The learner can select from a variety of information processing strategies depending on how he plans to use the information at a later time. Some tasks require the learner to focus on isolated bits of information and to store the presented information for a very brief period of time. Simple rehearsal processes may be adequate to fulfill the requirements of such a task. On the other extreme, some tasks require the learner to focus on broad, integrative principles and to recall the material months or even years later. Such a task would seem to require active transformational and coding processes which are, as yet, not well understood.

The learner's activities which intervene between the presentation of the instructional task and the student's performance on the criterion task are assumed to be adaptive behaviors which are influenced by the learner's perception of the desired criterion performance. We have called a class of student behaviors which intervene between instruction and criterion performance, student instrumental activities (Di Vesta, et al., 1970, p. 6). Instrumental activities are those activities which are learner initiated or directed, and mediate between the instructional demands and the learning outcomes. Covert forms of instrumental activities include silent rehearsal, evoking images, constructing mnemonics or verbal associates. The most common forms of overt instrumental activities are note taking and verbalizing. It is likely that the student's expectancy of when presented material will have to be retrieved will influence his instrumental activity which in turn will determine what can be remembered after given intervals of time. Note taking, verbalizing, imaging and other instrumental activities influence what and how much can be remembered at a later time.

The purpose of the present research is to determine the effects of note taking and expected recall interval on both immediate and delayed recall of instructional material. Berliner (1970) reviewed the experimental research on the effects of note taking in instructional settings and concluded that the evidence "is not favorable regarding the utility of note taking" (p. 2). Recent research by Di Vesta and Gray (1971) and Peters and Harris (1970) showed beneficial learning effects associated with note taking activities in experimental instructional settings. It is hypothesized that different expectancies will result in different note taking activities and that the test interval expectancy will also influence how much is remembered.

Method

Experimental Design

Eight different experimental conditions were created by completely crossing three two-level factors: (1) immediate test expectancy, delayed test expectancy; (2) immediate test, delayed test; (3) note taking, no note taking. The immediate test was given to the subjects immediately after they had read the assigned material; the delayed test was given one week later.

<u>Subjects</u>

The subjects were 164 students from the introductory educational psychology course at The Pennsylvania State University who received course credit for their participation in the experiment.

<u>Materials</u>

The instructional material was a five-page passage titled, "The Origin of Species," dealing with the principles governing the development of species. A multiple-choice, essay, and free recall test were used as criterion measures. The multiple-choice test consisted of sentences taken directly from the instructional passage with key words deleted and four alternatives from which to choose. The essay questions were rather specific, referring to the main ideas in the passage, e.g., "What are the two basic conditions ... necessary for the evolution of distinct species?" The free recall task consisted of the instruction to "Write down as much as you can remember about the passage, 'The Origin of Species'."

<u>Procedure</u>

Subjects were recruited and participated in groups of four to eight. Each of the subjects was assigned randomly to one of the eight experimental conditions. The subjects were seated at a large table which was partitioned into eight cubicles so that the subjects were separated from one another. In each cubicle was the instruction sheet appropriate to that subject's treatment condition and the "Origin of Species" reading material. For those subjects in the note taking conditions, a yellow pad and pencil were provided.

After the subjects were seated the experimenter gave brief instructions to the effect that they would have two minutes to read the instructions and thirty minutes to study the task material. After answering questions, if any, the experimenter left the room and observed the subjects through a one-way mirror.

Each subject read the instruction sheet which corresponded to the treatment group to which he was assigned. Each set of instructions was basically the same in format, each differing only with respect to the time at which the criterion test would be administered and whether or not note taking was permitted. The free recall and essay criterion tests were briefly described on the instruction sheet. The subjects' understanding of the key parts of the instructions was checked with three questions following the instructions. All subjects in the note taking conditions were informed that they would not be able to use their notes at the time of the test.

At the end of the 30 minute study period the experimenter re-entered the room and collected the instruction sheets, and any notes that the subjects had made. The subjects in the delayed test conditions were dismissed and instructed to return one week later for the test. The subjects in the immediate test condition were given the free recall and essay tests at this time. In the free recall test, the subjects were instructed to write down as much as they could remember in ten minutes. They were then given three essay questions and twelve minutes to answer them. After completing the essay test, they were told to return one week later for the second part of the experiment. Although the subjects in the immediate condition had been told they would receive an immediate test the request to return one week later came as no surprise to them because the experiment had been described as a two part experiment when they signed up for participation.

When the subjects returned one week later they were again seated in the same experimental room and given the same free recall and essay tests. For the subjects in the immediate test condition this was their second time taking these tests, but for subjects in the delayed test conditions this was the first and only time they would take these tests. After finishing the free recall and essay tests all subjects were given a 20-item multiple-choice test and 15 minutes in which to complete it. All time limits for tests were sufficient to provide no time restrictions on performance. After the multiple-choice test, they were given a brief post-experimental questionnaire which asked about the credibility of the experimental conditions and whether or not they had discussed the experiment with anyone.

Scoring

The score on the multiple-choice test was the number of correct responses. A scoring guide which listed the important parts of the answer to each essay question was used to score the essay responses. One point was awarded for each statement in the answer which coincided with the parts of the answer provided on the scoring guide. The free recall tests were scored for number of correct statements made. Each independent clause in the response on the free recall protocol was scored correct or incorrect and the total score represented the number of correct statements made.

Results

The means and standard deviations for the free recall and essay tests are given in Table 1. A 2 x 2 x 2 analysis of variance was carried out to test the simple and interactive effects of expected test interval, actual test interval, and note taking conditons. Separate analyses were done using the free recall and the essay test scores as dependent measures.

The analysis of scores on the free recall test indicated no significant main effects due to test interval expectancy, $\underline{F}$ (1,156) = .17, $\underline{p} > .05$, a significant main effect due to actual test interval, $\underline{F}$ (1,156) = 83.38, $\underline{p} < .01$, and no significant effect due to note taking, $\underline{F}$ (1,156) = .65. There was a significant interaction between the expected recall interval and actual recall interval factors, $\underline{F}$ (1,156) = 5.22, $\underline{p} < .05$. None of the other interactions were significant.

To further analyze the interaction effect, simple effects analyses were done on the expected recall interval factor. The mean for the immediate test was greater in the immediate test expectancy condition ($\bar{X} = 6.15$) than in the delayed test expectancy condition ($\bar{X} = 5.27$), but not significantly so, $\underline{F}$ (1,76) = 3.58, $\underline{p} > .05$. The mean for the delayed test was greater in the delayed test expectancy condition ($\bar{X} = 3.04$) than in the immediate test expectancy condition ($\bar{X} = 2.43$), but again the difference was not significant, $\underline{F}$ (1,76) = 1.72, $\underline{p} < .05$.

The analyses of scores on the essay test indicated no significant main effect due to test interval expectancy, $\underline{F}$ (1,156) = .86, $\underline{p} > .05$,

Table 1

Means and Standard Deviations for Free Recall and Essay Test Scores For Eight Experimental Groups

Expected Recall Interval	Actual Recall Interval	Test	No Notes Mean	No Notes SD	Notes Mean	Notes SD
Immediate	Immediate	Free Recall	6.45	2.87	5.84	2.34
		Essay	4.70	3.20	5.37	2.85
	Delayed	Free Recall	2.90	2.21	1.95	1.29
		Essay	2.81	1.78	1.86	1.81
Delayed	Immediate	Free Recall	4.74	2.10	5.80	2.19
		Essay	4.89	3.09	5.60	3.77
	Delayed	Free Recall	3.32	1.84	2.76	1.51
		Essay	2.86	2.03	2.90	1.87

a significant effect due to actual test interval, F (1,156) = 38.11, p < .01, and no significant effect due to note taking, F (1,156) = .08, p >.05. None of the interactions were significant.

Table 2 presents the multiple-choice test results. It should be noted that only a delayed multiple-choice test was given. Due to time the multiple-choice test could not be given as an immediate test. Therefore, the mean scores presented in Table 2 for the immediate test condition are the mean scores for those groups which took the free recall and essay test immediately. But all multiple-choice test scores were obtained one week after the study session.

A 2 x 2 x 2 analysis of variance was carried out on the multiple choice test results. The main effect due to test interval expectancy was not significant, F (1,156) = .66, p > .05; the main effect due to actual test interval was significant, F (1,156) = 10.75, p < .01; the main effect due to note taking was not significant, F (1,156) = .14, p > .05. The only significant interaction was between test interval expectancy and note taking, F (1,156) = 9.86, p < .01.

A test on the simple effects related to the significant interaction indicated that the mean score for the no notes condition ($\overline{X}$ = 9.30) was significantly higher than the mean score for the notes condition ($\overline{X}$ = 8.26, F (1,76) = 6.00, p < .01, in the immediate test expectancy condition. In the delayed test expectancy condition, the mean score for the no notes condition ($\overline{X}$ = 9.28) was greater than for the notes condition ($\overline{X}$ = 8.07) but not significantly so, F (1,76) = 3.70, p > .05.

In order to analyze the effects of test interval expectancy on note taking behavior, an analyses was carried out on the number of notes taken during study in the note taking conditions. The means and standard

Table 2

Means and Standard Deviations of Multiple-Choice Test Scores

For Eight Experimental Groups (Delayed Test Only)

Expected Recall Interval	Actual Recall Interval	Note Taking Condition: No Notes Mean	No Notes SD	Notes Mean	Notes SD
Immediate	Immediate	10.75	2.20	9.16	3.42
	Delayed	8.86	2.97	7.36	2.57
Delayed	Immediate	8.68	3.38	9.70	3.26
	Delayed	7.45	2.65	8.86	1.81

deviations for number of lines of notes taken in each of the four note taking experimental conditions is presented in Table 3. A 2 x 2 analysis of variance indicated a significant main effect due to test interval expectancy, $\underline{F}$ (1,76) = 19.6, $\underline{p} < .01$. The main effect due to the actual test interval was not significant, $\underline{F}$ (1,76) = 1.2, $\underline{p} > .05$, nor was the interaction between the two factors significant, $\underline{F}$ (1,76) = .28.

Discussion

The test interval expectancy factor did not emerge as a significant main effect in any of the three analyses. Note taking was influenced by the test interval expectancy, with subjects in the delayed test expectancy condition taking almost twice as many notes as in the immediate test expectancy condition. In a previous research study, (Weener, 1970) using the same instructional materials, subjects in the immediate expectancy condition scored significantly higher than the delayed expectancy subjects on an essay test. In that study also, the delayed expectancy subjects took almost twice as many notes as in the immediate expectancy condition. This latter finding must be interpreted in light of the fact that the subjects were instructed that they would not be able to use their notes later. Obviously, subjects do modify their instrumental activity depending on when they expect to recall the presented instructional material, but in the present research the increased note taking activity was not significantly associated with performance levels on any of the three dependent measures.

The interaction between the test interval expectancy factor and the note taking factor was significant for the multiple-choice test results ($\underline{p} < .01$), and approached significance in the analyses of the free

Table 3

Means and Standard Deviations for Number of Words of Notes Taken in Four Note Taking Conditions

Expected Recall Interval	Actual Recall Interval: Immediate Mean	Immediate SD	Delayed Mean	Delayed SD
Immediate	87.2	64.7	97.05	69.24
Delayed	155.4	90.17	183.8	86.3

recall results (F = 2.52, p = .12). When subjects were expecting an immediate test, those subjects who took notes did not do as well as those subjects who did not take notes. On a recognition task, such as the multiple-choice test, note taking apparently does not facilitate - and probably interferes with - the optimal learning strategies. When a person expects an immediate test, he probably scans as much of the material as possible with simple rote rehearsal being a dominant learning strategy. Note taking probably interferes with this activity and results in a decrement in recall performance when compared to the note taking conditions. When a person expects a test one week later with no opportunity for review in the interval, he probably recodes and transforms the data in order to make the information more resistant to forgetting. Note taking reinforces such instrumental activity. It can be argued that the function of most note taking activity is to recode and transform the data to make it more permanently imbedded in a person's cognitive structure. It is out of just such an awareness that subjects take many more notes when they expect a test several days or weeks in the future than when they expect a test within minutes after the completion of a task. In summary, note taking seems to interfere with the learning strategies employed by a person preparing for an immediate test, but seems to facilitate the learning strategies employed by a person preparing for a test in the future.

The interaction between expected test interval and actual test interval supports the hypothesis that subjects do modify their instrumental activities depending on when they expect to take a test on the material. Subjects who expected an immediate test did better on the free recall immediate test than those subjects who expected a delayed test; subjects who expected a delayed test did better on the delayed test than those subjects who expected an immediate test. The instrumental activities facilitative of long term storage are not necessarily facilitative of "short time" storage, and vice versa, for the instrumental activities facilitative of short-time storage.

Note taking did not emerge as a significant main effect in any of the three analyses. This is contrary to the positive effects of note taking which were reported by Di Vesta and Gray (1971) and Peters and Harris (1970). The important difference between the present research and those two studies is that written stimulus materials were used in the present study and orally presented information was used in the other two studies. Note taking is apparently not as important an instrumental activity when the stimulus material is continually available for rereading as when the stimulus material is available for the duration of the subject's auditory memory span. If the material is presented orally and not written down or actively assimilated during the brief period that it can be held in short term memory, the material is irretrievably lost. In this situation, note taking becomes an important activity to facilitate the recoding and transforming of the data as well as to provide an external storage for later reference.

References

Berliner, K. C. Effects of test like events and note taking on learning from lecture instruction. Amherst, Mass.: Technical Report, Center for Educational Research, March, 1970.

Di Vesta, F. J., & Gray, G. S. Listening and note taking. Journal of Educational Psychology, 1971, in press.

Di Vesta, F. J., Sanders, N. M. Schultz, C. B., & Weener, P. D. Instructional strategies: Multivariable studies of psychological processes related to instruction. Semi-annual Report, July, 1970.

Peters, K., & Harris, R. C. Note taking and review in recognition learning. In F. J. Di Vesta, et al., Instructional Strategies, Annual Report, June, 1970.

Weener, P. The effects of recall mode and recall interval expectancies on note taking and recall. In F. J. Di Vesta, et al., Instructional Strategies, Seni-annual Report, December, 1970.

Experimenter Attitude Effect and Subject Performance*

Study Director: Samuel K. Rock, Jr.

Advisors: Paul D. Weener and Nicholas M. Sanders

Technical Problem

Positive and negative experimenter attitudes were hypothesized to have demand characteristic functions (Orne, 1962) affecting subject performance on an experimental task. In particular, positive attitude was assumed to facilitate a subject's involvement in a learning task, while negative attitude was assumed to have a debilitating effect. It was reasoned that experimenter attitude would have a greater effect when subjects were given more study time and when subjects were tested on material directly related to specific study instructions (intentional material) than on incidental material.

General Methodology

In a laboratory setting, subjects were individually administered in sequence, a digit span task, a five-page prose passage learning task, and two tests, a free-recall test and a cued-recall test.

Half of the subjects were administered the tasks with the experimenter exhibiting a positive attitude, and the other half were

---

* This study was conducted as a Master's thesis in the Department of Educational Psychology and was supported, in part, under the present contract.

PRECEDING PAGE BLANK

administered the tasks with the experimenter exhibiting a negative attitude. The digit span task was used solely to provide extended exposure to the attitudinal behavior of the experimenter, and thus was not scored. In each attitude condition, half the subjects were allowed twice the length of time to study the prose passage as were the other subjects.

All subjects were instructed, prior to study of the passage, to pay particular attention to certain of its contents. Thus, the passage dealt with the lives of three educational psychologists and the subjects were directed to learn six specific points about each of their lives. This constituted the intentional task and responses were scored accordingly. Other information that subjects learned from the task was considered to be incidental learning for which a second score was obtained.

Technical Results

The results were essentially the same for both dependent measures, i.e., the free-recall test and the cued-recall test. The attitude of the experimenter had no overall effect on recall. However, the intentional learning scores were higher when the experimenter exhibited a positive attitude than when he exhibited a negative attitude. In contrast, incidental learning was lower in the positive attitude condition than in the negative attitude condition.

Those subjects who were allowed extra time for study retained more of the information than those subjects who were allowed less time. The data indicated that the extra time was used to learn more of the material the subjects were directed to learn (intentional learning).

The predicted effects of the interaction between experimenter attitude and rehearsal time on performance was not supported.

Educational Implications

At the present state of the science, the results of this study are more appropriately applied to educational research experiments than to the broader areas of application to the classroom. The results imply that experimenter attitude does influence subject performance on an experimental task(s), and has the potential of being an important variable, the effects of which may be unnoticed, in educational research. If the effects of the experimenter's attitude is not carefully controlled it may be a source of contamination of experimental treatments. Such confounding of effects might occur especially if the experimenter had a positive attitude at the beginning of an experiment and changed his attitude to a negative one as the experiment progressed, in which case subjects at the beginning of the experiment would be exposed to a different set of stimuli than the subjects at the end of the experiment. The consequence is differential performance at each phase unrelated to the hypothesized treatments, and the ultimate hiding of the differences in performance at each phase. Standardization of experimental procedures is vital but can be attained through such devices as tape-recorders, computers, and videotape to insure objectivity. While these devices may not eliminate the demand characteristics entirely, they can hold the effect constant over all treatments and subjects.

Teacher attitude may have a similar effect in the classroom but until the appropriate research is conducted any generalizations from the present study must be cautiously qualified. Nevertheless, it seems reasonable to hypothesize that the teacher's attitude will influence the student's attention to the specific material presented and thus influence his level of performance when tested on that material.

Implications for Further Research

Prior to extensions of research on this effect to school settings, other studies should be conducted with groups, rather than with individuals. If the effect does not occur in experimental groups then generalization to a classroom setting might be very limited. Further research should also be conducted with other dependent measures to determine the generalizability of effects obtained in the present study with meaningful prose. Should both these lines of investigation prove significant then classroom studies would definitely be indicated.

References

Orne, M. T. On the social psychology of the psychological experiment: demand characteristics. *American Psychologist*, 1962, *11*, 776-783.

APPENDIX

The research which has been supported in full or in part by the Advanced Research Projects Agency (ARPA Order No. 1269) through the Office of Naval Research is listed below. These include technical reports, theses, and research reports which have been published or are currently in press.

Published or In-Press

Di Vesta, F. J. An evolving theory of instruction. Educational Technology, 1971, in press.

Di Vesta, F. J., & Gray, G. S. Listening and note-taking. Journal of Educational Psychology, 1971, in press.

Di Vesta, F. J., Ingersoll, G. M., & Sunshine, P. M. A factor analysis of imagery tests. Journal of Verbal Learning and Verbal Behavior, 1971, in press.

Di Vesta, F. J., & Rickards, J. P. The effects of labeling and articulation on the attainment of concrete, abstract, and number concepts. Journal of Experimental Psychology, 1971, 88, 41-49.

Di Vesta, F. J., & Ross, S. M. Imagery ability, abstractness and word order as variables in recall of adjectives and nouns. Journal of Verbal Learning and Verbal Behavior, 1971, in press.

Peters, D. L., & Messier, V. The effect of question sequence in objective test performance. The Alberta Journal of Educational Research, 1970, 16, 253-265.

Sanders, N. M., Di Vesta, F. J., & Gray, G. S. The effects of concept-instance sequence as a function of stage of learning and learner strategy. Journal of Educational Psychology, 1971, in press.

Schultz, C. B., & Di Vesta, F. J. The effects of expert endorsement of beliefs on problem-solving behavior of high and low dogmatics. Journal of Educational Psychology, 1971, in press.

Schultz, C. B., & Di Vesta, F. J. The effects of passage organization and note-taking on the selection of clustering strategies and on recall of textual materials. Journal of Educational Psychology, 1971, in press.

Technical Reports

Di Vesta, F. J. Recognition and recall by high and low imagers of stimuli presented as words and as pictures. (July, 1971, Part II)

Di Vesta, F. J. Verbal and imaginal processing in learning and transfer by high and low imagers. (July, 1971, Part II)

Di Vesta, F. J., & Gray, S. G. Listening and note-taking II: Thematic content, note-taking, and length of listening review intervals as variables influencing immediate and delayed recall. (July, 1971, Part II)

Di Vesta, F. J., Gray, S. G., Ingersoll, G. M., & Sunshine, P. M. The effects of rated vividness and imagery of learning materials on learning and recall of high and low imagers. (July, 1971, Part II)

Di Vesta, F. J., & Ross, S. M. Contextual cues and cognitive structure in the storage and retrieval of information. (July, 1970)

Di Vesta, F. J., Schultz, C. B., & Dangel, T. R. The effects of congruence between passage organization and an imposed clustering strategy on recall. (July, 1971, Part II)

Di Vesta, F. J., & Sunshine, P. M. The retrieval of abstract and concrete materials by high and low imagers employing imaginal and verbal mediation with abstract and concrete mnemonic aids. (July, 1971, Part II)

Gray, G. S., & Di Vesta, F. J. The effects of search strategies on the learning of concept-attitudes. (January, 1971)

Peters, D. L. The effects of note-taking and rate of presentation on short-term objective test performance. (January, 1971)

Peters. D. L., & Harris, C. Note-taking and review in reception learning. (July, 1970)

Sanders, N. M. The effects of dogmatism on learning and transfer in concept-based and rote-based classification tasks. (January, 1970)

Sanders, N. M. The effects of studying together and grading procedures on recall of subject matter. (July, 1970)

Sanders, N. M., & Tzeng, O. The relationship between rote and conceptual learning. (January, 1971)

Sanders, N. M., & Tzeng, O. Type-of-test expectancy effects on learning of word lists and prose passages. (July, 1971, Part II)

Sanders, N. M., & Weener, P. D. Reliabilities of six personality measures used in the instructional strategies research project. (July, 1970)

Schultz, C. B. Recitation strategies: The effects of rates and schedules of verbal responses on retention. (July, 1970)

Schultz, C. B., & Dangel, T. R. Recitation strategies II: The effects of a learner-sustaining climate and encoding on the retention of facilitators and debilitators. (July, 1971, Part II)

Weener, P. D. Achievement anxiety and performance on the remote associates test. (January, 1970)

Weener, P. D. Small group verbal presentations, anxiety level, and learning in instructional settings. (July, 1970)

Weener, P. D. The effects of recall mode and recall interval expectancies on note-taking and recall. (January, 1971)

Weener, P. D., & Rock, S. K. The effects of recall interval expectancy and note-taking on immediate and delayed retention. (July, 1971, Part II)

Weener, P. D., & Tzeng, O. The effects of subjective organization instructions and verbal creativity on the recall of random and organized lists. (July, 1971, Part II)

Theoretical and Review Articles

Di Vesta, F. J. A theory of instructional strategy. (July, 1971, Part I)

Di Vesta, F. J. Instructional strategies and learner transformations. (July, 1971, Part I)

Sanders, N. M. Strategies for orienting students to educational objectives. (July, 1971, Part I)

Schultz, C. B. Cognitive stimulation in instructional strategies. (July, 1971, Part I)

Weener, P. D. Student instrumental activities and instructional strategies. (July, 1971, Part I)

Theses

Ingersoll, G. M. The effects of presentation modalities and modality preferences on learning and recall. Ph.D., 1971.

Marlow, E. S., III. Need for achievement, fear of failure, and attribution processes in goal setting. Ph.D. (Tentative title. Approximate completion date: September, 1971).

Rickards, J. P. Spacing, positioning, conceptual level, and type of statement as variables in the free recall of prose material. Ph.D. (Tentative title. Approximate completion date: September, 1971).

Rock, S. K. Experimenter attitude effect and subject performance. M.S. (Tentative title. Approximate completion date: September, 1971).

Ross, S. M. The effect of context modality on acquisition and transfer by imagers and non-imagers. M.S. (Tentative title. Approximate completion date: December, 1971).

Schultz, C. B. The effects of uncertainty, confidence, and individual differences on the initiation and direction of information-seeking behaviors. Ph.D., 1970.

Silvestro, J. R. Satiation of divergent and convergent thinking and its effect on the need for novelty. M.S., 1970.